THE PROCEEDINGS OF
THE PHYSICAL SOCIETY
Section B

Vol. 63, Part 7 1 July 1950 No. 367 B

The Sliding Surface*

By G. I. FINCH

34th Guthrie Lecture, delivered 3rd March 1950

§ 1. INTRODUCTION

Life is possible because some assemblages of matter can store sunlight as chemical energy. Other organisms, unable to do this directly, absorb these stores to supply their own energy needs; and some, like man, prey on both to live. Man alone supplements the body's production of physical energy by using inanimate devices to unlock the stores of energy latent in the remnants of past life. The key is the machine, which works because smooth surfaces can slide over each other. The sliding surface is, therefore, fundamental to our modern way of life.

The sliding surface was known long before the machine; away back in antiquity man had used it to ease the burden on his back. He had also learnt the value of a generous supply of some slippery substance. So when the machine did arrive the engineer saw to it that its sliding surfaces were lubricated; so well, indeed, that they were kept apart by an oil layer many hundreds or even thousands of molecules thick. Nowadays, however, sliding surfaces are usually so heavily pressed together that the lubricant film cannot be more than a few molecules thick, and is often penetrated or disrupted so that the sliding surfaces touch in places.

Thick-layer, or, as it is generally known, hydrodynamic lubrication (Reynolds 1886) favours easy sliding without wear other than that due to corrosion and erosion, but with boundary (or molecular film) conditions † of lubrication, where the sliding surfaces often come into contact, suppression of wear becomes the main problem. Thus, in hydrodynamic lubrication chief interest is directed to the lubricant, but in boundary lubrication it is the sliding surface which claims first attention. What causes rubbing surfaces to wear and how do they wear? What peculiarity in the surface helps to reduce wear and how does it work? How do the surfaces slide when they touch each other? What is it about the surfaces that helps them to slide and how does it help? We must first seek the answers to questions like these before turning to the study of the lubricant boundary film.

As far back as the eighteenth century Coulomb (1781) thought that resistance to sliding was due to the interlocking of surface asperities, and that the fracture of these during sliding was responsible for wear. Ewing (1912) and Hardy and

---

* This Lecture will also be published in Section A in August.

† The term 'boundary conditions' was first used by Reynolds.

Hardy (1919), on the other hand, believed that the interplay of surface cohesive forces was the dominant factor. While Coulomb's view may once have served a useful purpose, it is completely at variance with modern engineering experience. Furthermore, the results of critical surface studies by Beilby (1921), Hardy (1936), Finch (1937), Bowden and Tabor (1939) and Parker and Hatch (1950) do not support Coulomb's hypothesis but are readily interpreted in terms of surface cohesive forces.

The forces which produce cohesion within a block of metal must also act at the surface of contact between two blocks; they should, therefore, resist stress as if the blocks were one. There are, indeed, facts which support this conclusion. Slip gauges left 'wrung' together eventually seize and are damaged when separated. Hardy and Hardy (1919) found that some solids seized together on contact. Fifteen years ago I showed—and it must have been known long before then—that mica cleavage surfaces could adhere firmly when brought together (Demonstration). But experience also shows that surfaces can, and often do, slide easily over each other, even in the absence of lubricating oil.

Although a solid exerts a cohesive force on an atom beyond its surface, the intensity of the force decreases rapidly with distance, to become negligible a few atomic radii away. Contact between two surfaces, therefore, implies approach to within this distance. It is possible, by cleaving certain mineral crystals, to produce atomically smooth plane surfaces of considerable area, but otherwise such a degree of smoothness cannot in general be achieved or even approached. Hence, when two nominally plane surfaces formed, let us suppose, on two blocks of the same metal are placed together, contact can occur only at three points. If the surfaces are now loaded, stress at the points of contact must exceed the elastic limit. Plastic flow will then enlarge the points of contact to areas, and fresh points of contact expanding to still more areas will be established, until the load is borne without further deformation of the material.* A considerable effort should now be required to pull or shear the blocks apart because, at the areas of contact, they should have become united. Experiment, however, gives a different result. Thus, mica cleavage surfaces which have been exposed to the atmosphere for some time, or momentarily immersed in water, will no longer adhere, no matter how hard they are forced together (Demonstration). If, instead of being 'wrung', two slip gauges are merely pressed together, they do not stick but fall apart when the pressure is released (Demonstration). Mechanically or anodically polished surfaces of copper, iron, silver, tin and other metals behave similarly, provided the pressure does not cause general plastic flow at the surfaces (Demonstration). An atomic degree of smoothness, superimposed, it is true, on a large-scale waviness of small amplitude, may be imparted to a face of a copper single-crystal by anodic polishing; nevertheless, a degree of pressure short of that causing permanent distortion will not even partially weld two such surfaces together (Demonstration). In this case the undimmed clarity of the electron diffraction pattern shows that any adsorbed layer cannot be more than two or three atoms thick (Figure 1).†

---

* I frame these conclusions on the basis of Amontons' (1699) experimental law according to which the coefficient of friction is independent of the area of the sliding surfaces. The validity of this law has often been confirmed. The reasons for apparent exceptions are well understood; they do not affect the issue.

† Figures 1 and 3–20 are photographs and are printed together opposite p. 472.

It may therefore be inferred that true contact is not as a rule established when even atomically smooth solid surfaces are pressed together, but that they are kept apart by adsorbed surface layers which, while satisfying the cohesive forces at the surfaces, exert little or no attraction towards each other. The remarkable resistance of these layers to penetration appears to be independent of their nature (if it is assumed that this varies with the metal) and determined mainly by the physical properties of the metal, general plastic flow of which might be expected to rupture such a layer by extending its surface, and thus allowing welding to occur through direct metal-to-metal contact. It will be evident that this adsorbed layer is of vital importance. Without it, a pair of surfaces at rest under load would seize together over an area proportional to the imposed load. It cannot be supposed that welds are formed only to be broken when, on release of pressure, the surfaces are forced apart by a recovery of the metal about those areas where the elastic limit has not been exceeded; for the metal in such welds, being itself compressed to the elastic limit, would have the greater range of recovery on release of the pressure.

When the surfaces are not only pressed together but forced to slide past each other, the effectiveness of the adsorbed layers in preventing metal-to-metal contact varies greatly according to the nature of the metal. Anodically polished single-crystal copper or aluminium surfaces at first slide then suddenly lock firmly together, even with light loading (Demonstration). Similarly prepared polycrystalline surfaces of these metals also seize quickly, though not quite so readily as single crystals (Demonstration); the corresponding mechanically polished surfaces slide more easily and are much more resistant to seizure (Demonstration). Anodically polished polycrystalline silver or iron single-crystal pairs slide freely and must be heavily loaded before signs of seizure appear (Demonstration), and similar experiments with combinations of anodically polished single-crystal and polycrystalline surfaces of various metals reveal further wide differences in the ease or otherwise with which sliding and particularly seizure take place. But whether this is to be attributed to the physical properties of the metal, as in the case of the loaded surface pair at rest, or whether the nature of the adsorbed layer is also significant, remains to be seen.

The existence of these layers has long been known. Faraday (1838) and De la Rive (1839) recognized their importance in heterogeneous catalysis, and they have since been much studied by the chemist. The physicist's interest seems to have been first fully aroused by their effect on thermionic and photoelectric emission (Langmuir 1916). The engineer, so long accustomed to hydrodynamic lubrication, has been slower to appreciate their value.

The layer is difficult to remove from a metal surface. It can be done by scraping or evaporating *in vacuo* and so exposing a fresh surface, or by heating to a high temperature *in vacuo* or in the presence of atomic hydrogen; but the layer re-forms spontaneously and rapidly, even at room temperature, on exposure to air. These facts suggest that the adsorption is chemical rather than physical, and that the layer is an oxide of the metal. This has been amply confirmed by electron diffraction.

Although oxidation is the immediate result of the exposure of a fresh metal surface to the atmosphere, the microbalance reveals that a moisture film of considerable thickness (as much as 80 molecular layers and possibly more) may be adsorbed on the oxide surface. When fresh, this layer is as a rule not strongly held and is removed by the electron beam in the evacuated diffraction camera.

§ 2. THE SLIDING ELEMENT

It is clear that the sliding element of the modern machine must be a highly complex system. Actually its simplest form comprises at least nine distinct components (Figure 2). These are (i) and (ii) the metallic backings supporting (iii) and (iv) the pair of metal surface layers or so-called bearing surfaces which, in turn, carry (v) and (vi) their appropriate oxide layers each with its (vii) and (viii) adsorbed boundary layer of lubricant, and (ix) an oil film of varying thickness held between the boundary layers and more or less charged with debris, moisture and gas bubbles. The components in each pair may differ widely in nature and properties, even in the case of the two lubricant boundary layers. Furthermore, a change in any one component may also affect the behaviour of the others. It is this astonishing complexity in what at first sight appears to be a simple system that explains why, in the past, research has been for the most part directed towards the special circumstance of practice, rather than to a fundamental study from which broad generalizations might be drawn. Clearly, the physicist's approach to the problem must be to break down the sliding element into its simplest components and to examine each in turn, first by itself and then in relation to the others.

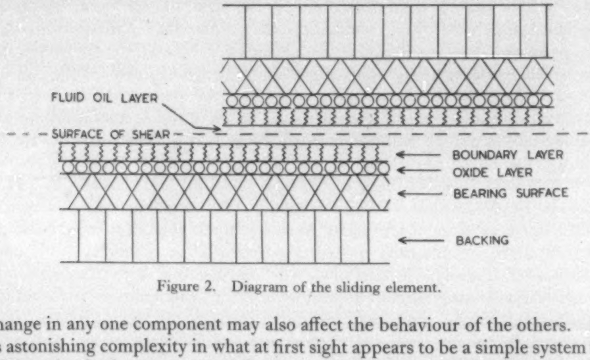


Figure 2. Diagram of the sliding element.

§ 3. THE BACKING

The backing is rigidly secured, directly or indirectly, to an appropriate member of the machine. Its function is to support the bearing surface so that its motion is constrained to the desired degree or degrees of freedom with respect to the other sliding member. In the case of a journal bearing, the force tending to rotate the backing in its housing may be considerable and is given by the product of the coefficient of friction between the sliding members and the load. Since the polar distribution of the load may vary periodically between wide extremes, particularly in the case of the internal combustion engine, the backing tends to oscillate in the housing. This may lead to serious frettage corrosion and loss of fit.

Backing and bearing surface may be composed of the same metal, as in the shaft of a journal bearing; or the metals may differ, as in the liner of the housing within which the shaft rotates. As the properties of the backing may affect the behaviour of the bearing surface, it will be convenient to defer its further discussion.

§4. THE BEARING SURFACE

The bearing surface is developed in practice as a plane or surface of revolution by abrasion processes such as cutting, grinding, honing and lapping. These operations reduce the crystal size of the metal at the surface, and often to a considerable depth below it (Figure 3). This 'work-hardening' can seriously impair the efficiency of the bearing surface, and for this reason the more severe operations of cutting and grinding should be followed by annealing to restore the uniformity of crystal size and relieve stress (Finch 1945). The texture of the finished surface is, in terms of atomic dimensions, exceedingly rough and made up of rows of steep hills and valleys, each with smaller irregularities and lying roughly parallel to the direction of cutting. The ridge crests are seen by electron diffraction to be crystalline (Figure 4).

General experience has shown that the fresh bearing surface must not be subjected to full loads and sliding speeds until it has been 'run-in' under light loads with copious supplies of lubricant. If the protuberances are not gradually reduced, as they are by 'running-in', the crystal structure will be broken up to such a depth below the surface as permanently to impair the sliding and resistance to wear of the surface. The effect of 'running-in' may be explained briefly as follows. Consider, first, bearing surfaces of similar metal. When, through breakdown of the oil and oxide layers, metal-to-metal contact is established between opposing high spots of the bearing surfaces, a weld is formed. As sliding motion proceeds the weld is stressed and eventually ruptured. If the crystal structure in the surfaces has not previously been broken up and therefore hardened, the weld will shear at or near one of the surfaces. On the other hand, if, owing to previous unduly severe conditions of machining or sliding, work-hardening has occurred to some depth below the surface, the weld, instead of rupturing virtually in the plane of sliding, will continue to adhere to one surface and pull out work-hardened metal to a considerable depth from the other (Figure 5). The protruding metal will then score and break up the crystal structure in this surface until it has been worn away; or it may break off and roll between the surfaces and damage them by crushing (Figure 6).

It is now known that sliding causes a profound and important change, not only in the texture but also in the structure of the metal surface. At first, rubbing of the sharper asperities on the opposing fresh bearing surfaces results largely in wear and the production of debris. As the surface becomes smoother, however, wear decreases and much of the metal removed from the hills 'flows', as Beilby (1921) puts it, into the valleys. Thus 'running-in' levels off the mountain ridges and fills the valleys to leave a texture like undulating hills. This smoothing is accompanied by a remarkable change in structure (Finch, Quarrell and Wilman 1935). Not only are the crystals broken up to a depth below the surface depending on the severity of 'running-in', but the surface layer structure becomes amorphous (Figure 7), as if 'flowing' consisted of the transfer of molten metal from one part of the surface to another where, having escaped the polishing action, it freezes too rapidly to crystallize. The thickness of this so-called Beilby layer depends, like that of the work-hardened crystalline zone beneath, on the severity of the 'running-in' conditions. Layers of only about fifteen atoms thick have been produced (Hopkins 1935) and others about 100 atoms thick have been successfully detached from their substrates (Cochrane 1938); but a vigorous polishing results in a much thicker layer (Finch, Quarrell and Roebuck 1934).

The following experiments will show how great is the protection afforded to the oxide layer by the support of even a very thin Beilby layer. Severe welding and extensive damage occurs when two anodically polished copper single-crystal surfaces are slid past each other (Figure 8) (Demonstration). The behaviour of a similar pair of surfaces, on one of which, however, a very thin Beilby layer has been formed by light polishing with rouge, is very different. Much heavier pressure is needed to cause welding, and on forcible separation of the surfaces metal is pulled out from the polished surface only to a shallow depth (Figure 9) (Demonstration). Thus the Beilby layer performs important functions in sliding. Its comparative smoothness favours a wide distribution of load where 'high spots' meet as the bearing surfaces slide past each other; it is more rigid than the underlying crystalline metal and thus better able to support the oxide layer and preserve it from injury; and a cold weld, with its tendency to pull out near the zone where the amorphous merges into the undisturbed crystalline structure, will do least damage if the Beilby layer is thin.

The relative amounts of metal worn away (and thus lost to the surface) and metal which is merely displaced to form the Beilby layer, are affected by the direction of the ridge and valley lines produced in machining in relation to that of sliding (Finch and Zahoorbux 1937). It has been found that the heaviest wear occurs when these lines in both bearing surfaces lie in the direction of sliding, and that wear is least when they are at right angles to the sliding direction. If the lines on one surface are parallel to the direction of sliding, while those on the other are at right angles to it, then wear of the former surface is high. The explanation is simple. When the lines in both surfaces are parallel to the sliding direction, the metal, severely stressed in localized contacts has farther to flow before it can escape and congeal beyond the pressure zone, and is thus more likely to form unattached detritus than when the ridges and valleys lie at right angles to the sliding. The importance of the 'run' of tooling tracks in relation to the sliding direction is now appreciated in engineering practice.

We have seen that when two nominally plane surfaces are pressed together the total cross-sectional area of the resulting intermetallic contact is proportional to the load. The force required to start sliding is therefore also proportional to the load; its value per unit load is known as the coefficient of static friction. To maintain sliding, work must be done. The necessary force is again proportional to the load, provided shear occurs at the interface and the strength of the metal in the contacts is unchanged. Under such conditions the dynamic and static coefficients of friction should be equal. Micrographic examination shows, however, that the cold welds formed on contact between, for example, single-crystal copper surfaces are strain-hardened as sliding proceeds, and also that the surrounding metal is similarly affected to a considerable depth (Figures 5 and 8). Hence, instead of the welds shearing at the interface, metal is pulled out bodily from below one surface to be dragged along and ruck and score the other (Figure 5). Under such conditions the dynamic coefficient exceeds the static. In normal practical conditions of sliding, of course, the reverse is the case, as the oxide and lubricant films together greatly limit intermetallic contact during sliding and may even in some cases inhibit it. Nevertheless, as a rule, spasmodic intermetallic contact does occur, and its contribution to the total resistance to sliding may be called intermetallic friction. Its most undesirable feature is that it causes wear.

A sliding element in which the intermetallic component of the total friction is considerable is therefore unsatisfactory.

So far we have considered bearing surface pairs of the same metal. In practice, however, where sliding speeds are high, this combination is avoided except in the case of cast iron. Otherwise the engineer's empirical rule is that the mating bearing surfaces should consist of metals of widely differing melting points. It can, indeed, be inferred from Bowden and Ridler's (1936) measurements of the temperatures attained at metal contacts in sliding that localized fusion followed by hot welding should occur more readily with like than with unlike metals. The matter, however, goes deeper than this. Solid mercury and gold would make an unsatisfactory pair of bearing surfaces.

Now, we have seen that in addition to hot welding due to melting of rubbing high spots on both bearing surfaces, cold welding also occurs on contact of the solid metals. Indeed, under the conditions of the experiments I showed you, the friction and surface damage was entirely due to cold welding and its after effects. Even with like metals, it is only under severe sliding conditions and more particularly at high speeds that hot welding becomes important. It will be evident, therefore, that in selecting metals for a bearing surface pair, one criterion is the intensity of the cohesive force across the interface of intermetallic contact, rather than disparity in their melting points. If this interfacial cohesion is less than that within the weaker of the two metals, intermetallic contacts will shear in the interface. The other criterion, which concerns the nature of the oxide layer, will be discussed later.

The problem of estimating this cohesive force from a knowledge of the structures of the outer electron shells of the atoms of which the contacting surfaces are composed has not been solved. However, a study of the coordination relationships between epitaxial metallic deposits and metallic single-crystal substrates is now being carried out in my laboratory (Finch and Sun 1936, Finch and Williams 1937, Finch, Wilman and Yang 1947); this seems to offer a promising line of experimental approach to the problem. In the meantime, as a rough and ready rule, we may take it that the more rigid the metals of a pair and the higher their melting points, the less will be the tendency to form *cold* welds.*

It would, I think, be fair to say that, as in the development of most aspects of the sliding surface, the choice of metals for bearing surface pairs has been empirical. The modern journal bearing practice is to use pairs in which one metal is hard and has a high melting point and the other is relatively soft and fusible. In the sliding of such combinations new phenomena appear.

Bowden and Hughes (1937) showed that when certain solids were rubbed together it was not the harder surface which caused the softer to flow and thus be polished, but that the more fusible surface was polished by the more refractory one, irrespective of their relative hardnesses at room temperature. In practice, however, sliding surfaces may well behave differently. For example, a type of journal bearing in common use consists of a superficially hardened alloy steel shaft rotating against a bearing surface of a rather soft tin-base alloy melting at about 250° c. Since the steel melts at roughly 1,400° c. the alloy surface becomes polished as one expects; what is surprising, however, is that the hardened surface of the shaft also flows and is polished as running-in of the bearing proceeds. Similar observations have been made in reciprocating sliding. Indeed, in all practical cases so far examined this mutual polishing of sliding metal surfaces has proved to be the rule, irrespective of the difference in hardness or fusibility of the bearing surface metals. The explanation is to be sought in the nature and properties, including the geometry, of the oxide layers and metals of the bearing surfaces, and will be discussed later.

---

* It would be virtually impossible to put together even two perfectly plane complementary single-crystal surfaces on a pair of blocks of the same metal in such a way as to restore the original atom coordination across the interface. The higher the mobility of the atoms, however, the greater is the chance of the atoms in the opposing faces moving into and being retained in sites of lowest potential energy. This is akin to a process of recrystallization whereby the two blocks eventually become welded together into a single crystal.

Quite apart from the reduction in welding, and hence in friction and wear, further advantages can accrue from the use of a relatively fusible soft alloy sliding on a hard refractory bearing surface. Owing to the comparative ease with which the softer metal is plastically deformed, it was long thought necessary to restrict the load intensity. The further disadvantage of the low ductility and proneness to fracture of the customary rather thick strip of soft metal was overcome, at least in part, by welding to a massive backing of a metal of more suitable mechanical properties, which was then secured in an appropriate housing. The resulting bearing, however, was large and massive in relation to the applied load. Hopkins and Palm (1932) realized that, by taking advantage of the cohesion between the bearing surface and the backing metals, the softer metal could be made to withstand a load intensity far greater than hitherto thought practicable. The underlying principle is simple. The ability of a soft body to resist deformation when compressed between two hard plane surfaces increases as its thickness is reduced by plastic flow until, long before the soft material can be totally extruded, the hard metals themselves begin to flow and deform. For example, it is impossible to force out completely even an oil film from between two plane surfaces, no matter how heavy the load or how rigid the metal may be. This is due to the reinforcing of the cohesion between the oil molecules by their adhesion to the surfaces. In fact, the thinner the film, the more do its mechanical properties approximate to those of the solid substrate. Similarly, whilst a bearing metal layer of the once customary thickness can support only a relatively low load intensity, a much thinner layer can carry a far heavier load without being permanently deformed. The thickness of this layer need be no greater than that corresponding to the maximum wear loss which can be tolerated before the fit becomes so slack as to render the bearing useless. For example, the 'big-end' crankshaft bearing of an internal combustion engine, which previously consisted of a 5/100 in. thick tin-base alloy fused on to a massive bronze backing, has now been replaced by a 5/1,000 in. thick layer on a thin steel strip, the weights of the liners being 218 and 60 gm. respectively (Exhibits).

The tin-base type of bearing metal alloy, though relatively soft, is too hard to be scratched with the finger nail. This form of alloy, though extremely wear-resisting, causes considerable wear of the hardened steel surface against which it slides. As we shall see later, this is due partly to the softness of the alloy and partly to the nature of its oxide layer. There is little evidence to support the view that embedded detritus particles are the cause. A further notable advance therefore was made by the recent introduction of bearing metals whose surfaces cause very little wear of the steel bearing surface. Lead and indium, the chief metals of this kind, are so soft that they and their alloys can be scratched with the finger nail.

Figure 1. Electron-diffraction pattern from a copper single-crystal electropolished cube face; beam along cube edge. The sharp Kikuchi lines and the narrow circular Laue zone show that the surface is atomically smooth and of quite remarkable lattice perfection. The clarity in detail near the shadow edge is such as to show that the oxide layer cannot be more than two or three molecules thick.

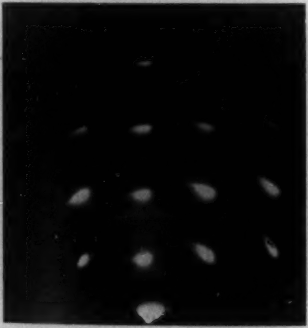


Figure 3. An iron single-crystal (110) face was lightly honed along a [1$\bar{1}$0] direction and then etched to expose the sub-surface damage. The pattern (beam along [001]) reveals that the original single-crystal structure has been distorted by a process of rotational slip (Wilman 1950) on (001) about a [001] axis. This slipping represents the initial stage in crystal break-up. Note that the fragments are still coherent across the planes of rotational slip.

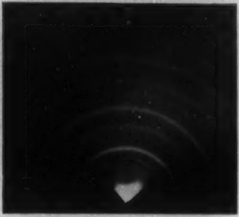


Figure 4. α-iron pattern from the freshly honed surface of an austenitic steel cylinder liner.

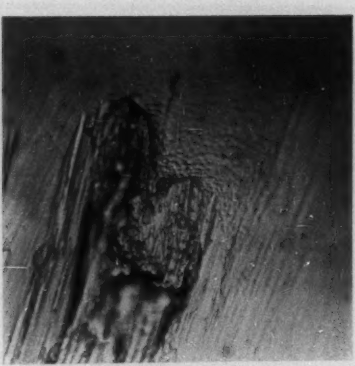


Figure 5. Scoring damage due to welding between a pair of copper single-crystal sliding surfaces. 450 ×.

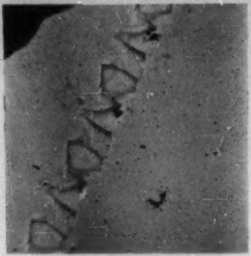


Figure 6. Track of a broken-off weld particle rolling on a copper single-crystal surface. 'Formvar' replica; gold shadowed. 10,000 ×.

Figure 7. Mechanically polished copper surface. The pattern consists of two ill-defined broad haloes. Patterns of this kind are obtained from glass, 'fused' quartz, cellulose, and from mechanically polished metal surfaces in general.

Figure 10. Pattern from an austenitic steel internal combustion engine liner after 120 hours' running. The halo pattern reveals the presence of amorphous iron and an amorphous iron oxide. The faint streak above the undeflected beam spot is due to a graphite film the crystals of which are orientated with their cleavage planes parallel to the liner surface.

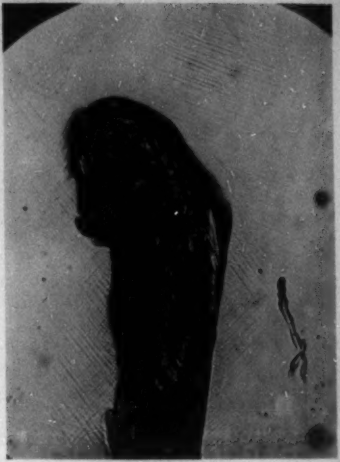


Figure 8. Surface damage resulting from weld formation between a pair of copper single-crystal surfaces. Note widespread regions of deformation bands formed ahead of and at the sides of the main damage area. 450 ×.

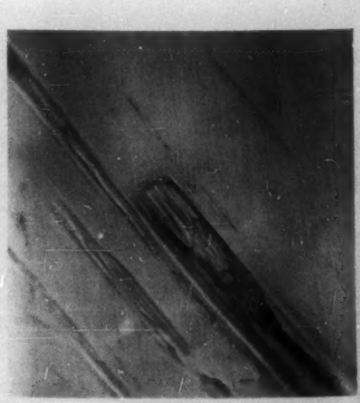


Figure 9. The weld damage is greatly restricted when one of the copper single-crystal surfaces is first lightly mechanically polished. The vertical lines are due to the polishing medium. 450 ×.

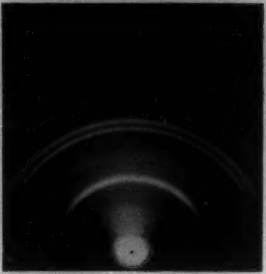


Figure 11. A scratched cast-iron piston-ring surface yields a pattern due to disorientated graphite crystals.

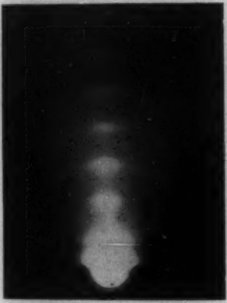


Figure 12. Light rubbing with the thumb suffices to orientate the graphite crystals with their cleavage planes parallel to the piston-ring surface. This pattern identifies the streak in Figure 10 as being due to graphite.

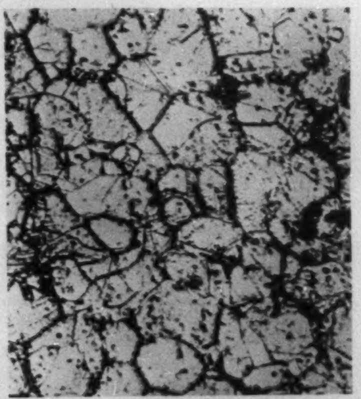


Figure 13. Polycrystalline copper surface after removing an oxide film. The crystal boundaries have been more attacked than the crystal surfaces.

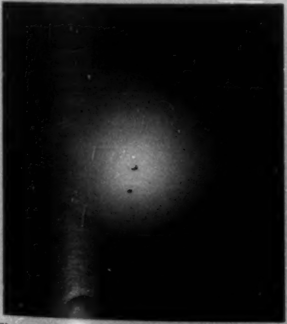


Figure 14. Amorphous iron oxide film removed from a mild steel surface.

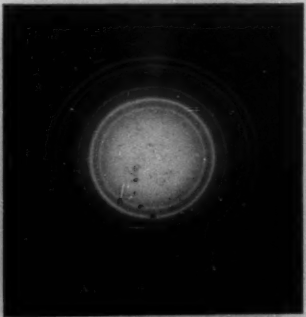


Figure 15. γ–Fe_2O_3 formed by heating the amorphous oxide film of Figure 14.

Figure 16. A copper single-crystal surface immediately after anodic polishing.

Figure 17. The same surface after 30 hours' exposure to air at room temperature becomes covered with an amorphous oxide layer which is thick enough (>40 Å.) to obscure the single-crystal pattern.

Figure 18. Cuprous oxide growing epitaxially on a heated macrocrystalline copper surface.

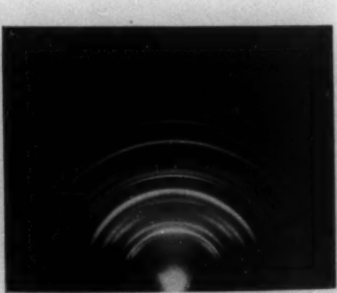


Figure 19. γ–Fe_2O_3 formed on a heated steel surface. The tendency is for needle-like crystals to form with [110] axes normal to the surface.

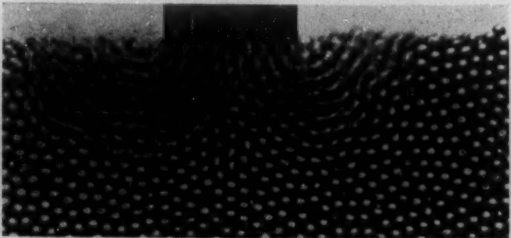


Figure 20. General movement of carbon shot during descent of piston. (*From* Andrade and Fox (1949).)

Nevertheless, when bonded in a sufficiently thin layer to a steel backing, they can support unusually heavy loading. One such backing-bearing surface combination consists of a hard steel strip to which is bonded a thin copper–bronze alloy strip on which a layer of lead is electrodeposited. A thin 'flash' of indium is electrodeposited and later thermally diffused into the lead. The copper–bronze strip is necessary as, owing to the differences in crystal structure and lattice constants, it is difficult to deposit lead electrolytically on steel in such a way as to ensure a satisfactorily close co-ordination, and therefore good cohesion, between the atoms across the interface. The total thickness of the lead–indium layer is 1/1,000 in. The rate of wear of both bearing surfaces is so small in relation to that of the other sliding surfaces of the engine that this thickness is ample (Exhibits).

An important question relating to bearing surfaces which has so far received less attention than it undoubtedly merits is that of the effect of sliding on the surface composition of alloys. In the few cases in which this has been examined experimentally it has been found that polishing increases the proportion of one or more of the constituents in the surface of the alloy, and hence also affects the composition of the oxide layer. Thus polishing a chromium steel alloy produces a surface which is either greatly enriched in chromium content or is even pure chromium covered by a layer of chromium oxide (Vernon, Wormwell and Nurse 1944). A similar effect has been observed in this laboratory with magnesium–aluminium alloys (De Brouckère 1945), α-brass and bronze (Bryant 1936) which become richer in magnesium, copper and tin respectively. Such changes may well be significant, not only because of their possible effect on the surface hardness and fusibility, but also because they must determine the nature of the oxide layer.

The demands made on the reciprocating bearing surfaces in the cylinder of the internal combustion engine are in many respects much more severe than in the journal type of bearing. Surface temperatures are high, particularly near the combustion zone, and the pressure distribution on the piston crown is seldom uniform, especially at the start of the expansion stroke. Further, the change from dynamic to static conditions at the end of each stroke not only imposes unusual demands on the lubricant layers but also results in an uneven load distribution between the sliding surfaces. Finally, the reaction products cause corrosion wear (Williams 1937).

Nominally, the bearing surfaces are those of the cylinder wall and piston rings, but the piston itself at times also rubs against the cylinder wall. For reasons of mechanical strength, lightness and thermal conductivity, the piston may consist of an aluminium alloy with copper and some magnesium, though in diesel engines a cast-iron skirt with an alloy steel crown is more usual. The cylinder bearing surface is a cast-iron, austenitic steel or other ferrous alloy, sometimes with a coating of electrodeposited chromium. The piston rings, however, are always made of cast iron. Alloys with better mechanical properties have been tried as piston-ring metal in low-compression engines but their adequate lubrication presents difficulties. The graphite in the cast-iron ring helps greatly in the suppression of wear of both cylinder and ring (Finch, Quarrell and Wilman 1935, Finch and Whitmore 1938). Electron diffraction has shown not only that the 'run-in' surfaces of the cylinder and cast-iron piston ring are covered with an amorphous oxide layer (Figure 10) but has also revealed the presence of an exceedingly thin adsorbed graphite layer (Figures 11 and 12), the flakes of which are orientated with their cleavage planes parallel to the surface. It is probable that this layer acts as a temporary lubricant and protection to the oxide layer when the oil boundary layer breaks down.*

The effect of direct sliding contact between the exceedingly hard corundum coating on the aluminium alloy piston and the oxide layer on the cylinder wall seems to be mitigated by the formation of magnesium spinel (Finch 1937). There is some evidence also that the rubbing of the corundum or spinel layers against the ferric oxide layer on the cylinder wall may lead to the formation of iron spinels (Thirsk and Whitmore 1940).

§ 5. THE OXIDE LAYER

In modern engineering practice, where frequent penetration of the lubricant boundary layer occurs, the part played by the oxide layers in facilitating sliding and preventing metal-to-metal contact is obviously of great importance. It is now known that most metals, if not indeed all, oxidize in dry air, however low the temperature. The primary oxide layer is formed very quickly, even at the temperature of liquid air, and then acts as a barrier against continued easy access of oxygen to the surface. Thus the subsequent rate of growth of the oxide layer decreases rapidly, and may become negligible at a thickness of the order of a few tens of atomic radii. Should further oxidation occur, it is very slow, even at room temperature, provided secondary effects, such as attack of the oxide films by adsorbed moisture, are avoided. Cabrera and Mott (1949) explain the mechanism of oxidation subsequent to the formation of the first oxygen layer in terms of the high contact potential gradient between the metal and the layer, which enables the metal ions to move through it without much help from thermal motion. There is, however, evidence which suggests that oxidation can also be due to oxygen gaining access to the metal through crystal boundaries. Thus, we have observed than an anodically polished single-crystal metal surface oxidizes more slowly than a similarly treated microcrystalline surface. Further, on removing with citric acid a relatively thick oxide film grown at 180° c. on anodically polished polycrystalline copper, it was seen that the crystal boundaries, barely visible before oxidation, had become strongly outlined and deepened (Figure 13).

The surface oxidation of metals is, in general, a highly energetic process. At low temperatures, therefore, the rapid arrival of oxygen molecules and the low mobility of its atoms over the surface may be expected to favour the growth of amorphous rather than crystalline layers. There are many examples of films having been grown on single-crystal substrates cooled to liquid air temperatures and, as far as I am aware, all have been found to be amorphous or so submicroscopically crystalline as to be indistinguishable from amorphous; yet the same substances deposited on the heated substrates have grown epitaxially. Even when a metal surface is oxidized at room temperature the oxide layer usually seems to be amorphous. This is certainly so in the cases of aluminium, chromium and iron (Figures 14 and 15); and recently Dr. Leu, working in my laboratory, has found that the oxide film formed on an anodically polished copper single-crystal face exposed to air at room temperature is also amorphous (Figures 16 and 17).

---

* It is commonly supposed that the crystals of layer-lattice structures such as mica, graphite and molybdenum sulphide, which are sometimes used as dry lubricants, act like orientated packs of playing cards. This cannot be so because, as I demonstrated earlier in this lecture, the layers can only be separated by cleaving in such a manner that the bonds between the layers are broken one by one. I also showed that, though freshly cleaved mica surfaces would adhere firmly on being first put together and could not be pulled apart by stress applied in a direction lying in the cleavage plane (i.e. attempting to rupture the bonds simultaneously), they lost this property on prolonged exposure to air or moisture. Thus the ability of layer-lattice cleavage surfaces to slide easily over each other is due to adsorbed layers. Layer-lattice crystals with their *c* axes parallel to, and crushed between, loaded sliding surfaces will be cleaved, but will also abrade the sliding surfaces. It follows that layer-lattice crystals in which the ratio of exposed cleavage plane area to thickness in the *c* direction is a maximum make the best form of solid lubricant. It is known from electron diffraction (Finch and Wilman 1936) and electron-microscopic studies (Rao, recent work in this laboratory) that the smaller the volume of a graphite crystal the higher is the value of the above ratio. A layer-lattice lubricant should, therefore, consist of crystals in the finest possible state of subdivision compatible with a maximum surface to thickness ratio. Sliding of the lubricated mechanism should not be relied on to break up the crystals, as this must result in wear.

On heating to above room temperature, amorphous oxide films differ widely in behaviour. Those on aluminium and chromium remain amorphous at temperatures up to about 600° c. and can grow in thickness only very slowly, if at all (Preston and Bircumshaw 1936, De Brouckère 1945). On copper, cuprous oxide grows epitaxially (Figure 18) and, at higher temperatures, becomes in turn superficially oxidized to cupric oxide (tenorite). These films, although coherent, are fairly easily permeable by both oxygen and copper ions, and thus do not give a high degree of protection against oxidation. Still less protection is given by the crystalline iron oxide films ($FeO \rightarrow Fe_3O_4 \rightarrow Fe_2O_3$) (Figure 19) formed under such conditions that the crystals have a marked tendency to grow outwards from the surface along an atom row of high atom population density. With other metals the crystalline oxide layer formed on heating is orientated in such a way that the crystals have in common a plane of high atom-population density parallel to the metal surface. They also cohere to each other and form a continuous film which hinders access of oxygen to the surface even when the metals are molten. Well-known examples are tin, zinc, antimony, bismuth, cadmium and lead. Oxide layers grown on heated metals when the rate of supply of oxygen is sufficiently reduced, are often epitaxial, i.e. the crystal orientation in the oxide layer is related to and determined by that of the metal surface. Such an oxide layer grown on a single crystal is itself virtually a single crystal. Epitaxial oxide growth can also occur during the oxidation of polycrystalline surfaces, even though the metal crystals have no common direction or directions of orientation. Then, however, the crystalline continuity of the deposited layer is further disturbed by the epitaxial reproduction of the irregularities which constitute the crystal boundaries.

Crystalline oxide layers on the sliding surface are probably of rare occurrence. Although a metal surface exposed momentarily through rubbing may be extremely hot, it is, as we know from the formation of the Beilby layer, disturbed and very rapidly cooled, so that there will be a tendency for any re-formed oxide layer to be amorphous. Many 'run-in' bearing surfaces have been examined by electron diffraction; their oxide layers have been found to be amorphous. The thickness of the amorphous films so far studied appears to be relatively uniform, so that the surface texture is a fair representation of that of the underlying metal. Indeed, the amorphous aluminium oxide layer has been used as a replica in the electron-microscopic study of the texture of aluminium surfaces.

The adhesion of the oxide film to its substrate can be expected to be good by reason of its nature, though pseudomorphic strain in the initial layers may not be without influence (Finch and Quarrell 1933, Cabrera and Mott 1949). As the film is amorphous, and therefore isotropic, it should be equally strong in all directions in the surface. Its cohesive strength in the case of iron and aluminium is shown by the way in which detached films, less than 100 A. thick, of amorphous ferric or aluminium oxide can be lifted off a water surface on a wide-meshed (80/inch) gauze without tearing, a process which also demonstrates their flexibility.

We may now examine how the oxide layer is so effective in preventing metal-to-metal contact when two surfaces are pressed together. When the surfaces first come together they make a three-point contact. This does not mean that three pairs of points meet. The probability is rather that three asperities of small radii of curvature will lodge on surfaces of large radii. Such a contact pair may be represented by Figure 20 (Andrade and Fox 1949). If pressure is now applied, the load will be distributed far more widely over the lower, flatter surface than over the point. Hence the point will flow and spread and, though in so doing its oxide layer may be disrupted and thus expose the metal surface, the distribution of pressure ensured by the other, largely intact, oxide layer reduces the loading intensity on the opposing metal surface. The oxide layers are even more effective when the initial point contacts occur between areas of low curvature. On loading, conditions then approximate to the case of the pressing together of two plane surfaces separated by adhesive layers. Not until extensive plastic flow takes place at both metal surfaces will their oxide films be broken and metallic contact be established. Since such contact cannot take place before the disruption of both oxide layers, it would be impossible to press two plane surfaces into metallic contact where one surface is soft and the other hard, unless the softer metal in flowing plastically were to shear the oxide layer off the hard surface. This occurs when, for example, a lead column is compressed between a pair of anodically polished plane surfaces of polycrystalline copper; after extensive flow, the lead begins to form welds with the copper (Demonstration).

We will now consider the behaviour of the oxide layers under normal sliding conditions, that is, after 'running-in' has prepared the surface for full loading. Where the lubricant boundary layers break down, the oxide layers must come into contact and slide over each other. Both films may then rupture, allowing metallic contact and cold or hot welding; or one film may be ruptured and the other survive; or both films may survive, with or without the help of a re-forming of the lubricant boundary layers. Wherever they are in sliding contact, heat will be liberated in the oxide layers.

The facts concerning the oxide layers which bear on the rôle they play in sliding may be briefly summarized as follows. They consist of extremely thin amorphous films strongly held to the metal (ionic binding). The melting points and hardnesses of some of the oxides and other products of possible interest in sliding surface practice are given in the following Table.

It will be evident from the table that the poor protection afforded by the oxide layers to single-crystal surfaces of copper or aluminium sliding under light loads at a low rate cannot be ascribed to either the refractoriness or hardness of the oxide layer. On the contrary, the oxide layer on the harder iron single-crystal surface gives a considerable measure of protection, although the fusibility and hardness of ferric oxide lie between those of cuprous oxide and aluminium oxide. We have also seen that a thin Beilby layer on one of a pair of copper single-crystal surfaces rubbing together somehow increases the resistance of its oxide layer to rupture.

Experiments with pairs of dissimilar metals lead to similar results. Thus a copper single crystal sliding on aluminium under a light load almost immediately seizes (Demonstration), and the welds break in the body of the aluminium; obviously both oxide layers are injured. On the other hand, when a copper single crystal slides over an iron single crystal under a much heavier load, no seizure occurs (Demonstration). Yet the cuprous oxide film is broken in many places

Table 1

Compound	M.P. (° c.)	Hardness (Moh)	M.P. of metal (° c.)
Fe_2O_3 (haematite)	1565	6½	1535
Fe_3O_4 (magnetite)	1538 (decomposes)	6½	
FeS_2 (pyrite)	1171	6½	
FeS (troilite)	1193	4½	
$FeCl_2$ (laurenite)	670	Soft	
FeOCl	—	—	
Fe_2P	1290	—	
Fe_3P	1100	—	
$(Fe, Ni)_9S_8$ (pentlandite)	878	3¼–4	
$FeCr_2O_4$ (chromite)	—	5½	
NiO (bunsenite)	2090	5½	1455
NiS (millerite)	797	3½	
$(Ni, Fe)S_2$ (bravoite)	~ 900	5½–6	
CoO	1935	—	1495
CoS (jaipurite)	>1165	—	
$(Co, Ni, Fe)S_2$ (hengleinite)	~1000	5	
Cr_2O_3	1990	—	1615
CrN	1500 (decomposes)	—	
CrF_3	1000	—	
Ag_2S (argentite)	825	2½	960
AgCl (cerargyrite)	455	1½	
MnO (manganosite)	1650	5½	1260
Mn_3O_4 (hausmannite)	1705	5½–6	
MnS_2 (hauerite)	—	4	
Al_2O_3 (corundum)	2050	9	660
MoO_3	700	Soft	2620
MoS_2 (molybdenite)	1185	Soft	
SnO_2 (cassiterite)	1127 (decomposes)	6–7	232
SnO	< 950 (decomposes)	—	
SnS	882	—	
TiO_2 (rutile)	2130	6–6½	1800
MgO (periclase)	2500–2800	5½–6	651
ZnO (zincite)	1670	4½	419
ZnS (sphalerite)	1049	4	
Cu_2O	1235	—	1083
CuO (tenorite)	1026 (decomposes to Cu_2O)	3–4	
CuCl (nantokite)	422	—	
Cu_2S (chalcocite)	1100	3	
CuS (covellite)	220 (decomposes)	—	
PbO_2 (plattnerite)	290 (decomposes)	5½	
PbO (litharge)	888	2	327
$PbCl_2 . 2PbO$ (mendipite)	693	3	
$PbO . PbSO_4$ (lanarkite)	977	2½	
$PbSO_4$ (anglesite)	1000 (decomposes)	3	
$PbFe_4O_7$ (plumboferrite)	1530	5	
CdO	1426	3	321
CdS (greenockite)	900 (sublimes)	3½	
ZrO_2 (zirconia)	>2950	6½	1900
Sb_2O_3	656	2½	630
Sb_2S_3 (stibnite)	550	2	
Bi_2O_3	820	4½	271
V_2O_3	1970	—	1720
InO_2	700 (sublimes)	—	155
Spinels (Mg, Fe, Ni, Cr, Mn, or $ZnAl_2O_4$)	1800	6–8	
Magnetites (Mg, Fe, Zn, or $NiFe_2O_4$)	1500	5½–6½	
Chromites (Mg or $FeCr_2O_4$)		5½	

Among other surface compounds of possible interest are the selenides, tellurides, arsenides, phosphates and borates.

and the structure of the underlying metal is disturbed; the iron oxide, however, survives and thus prevents metal-to-metal contact. The oxide layers on surfaces of metals of high rigidity, such as hardened alloy steels, resist rupture at low sliding speeds, even when heavily loaded. Examples are afforded by ball-bearing and gear-tooth surfaces in both of which, although the load intensity is high, the sliding speeds are low or very low. Such bearings, even without a lubricant, are for a time capable of carrying surprisingly heavy loads at low speeds without forming welds. It is evident, therefore, that the resistance of the oxide layer to injury during sliding must, irrespective of its nature, depend in the first place on the rigidity of the supporting metal.

As the severity of the sliding conditions increases, however, it becomes apparent that the behaviour of the oxide layer is determined not only by the supporting metal but also by its own properties. For example, when the bearing surfaces consist of the same hard metal, the frequency of intermetallic contact and welding increases when the sliding speed or load is raised. As we shall see later, however, a change in the chemical composition of the oxide layer alone may reduce or even eliminate intermetallic contact and welding.

It seems to me that these facts are best explained as follows. The ratio of the load to the area of contact is constant and a measure of the rigidity of the softer of the two metals of the opposing surfaces. Hence, with a given load, the loading intensity increases with the rigidity of the softer metal or with that of the metal common to both surfaces. Therefore, since heat is developed in the oxide layers wherever their surfaces are sliding in contact, the local temperature attained must, under comparable conditions, be higher with rigid metals than when one or both are soft. Further, these local temperatures must increase with the speed of sliding. High local temperatures have, indeed, been observed during the sliding of glass or quartz surface pairs (Bowden, Stone and Tudor 1947), and this suggests that the factor limiting the possible temperature rise is the fusion of the oxide layer. If the softening temperature of the oxide layer is higher than that of the metal, the latter will melt and thus no longer support the oxide layer, and the ensuing hot metal-to-metal contact will lead to the formation of a hot weld.

To test this view we may consider the hypoid gear which was introduced about 1930. Though the sliding speeds in this type of gear are low, the load on the bearing surfaces is unusually high, and excessive wear due to welding was common. Some years previously it had been found that wear in cutting tools could be reduced by adding finely divided sulphur to the cutting lubricant. The sulphur seemed to act by forming a protective film on the tool tip. When, therefore, the best of the then known gear lubricants, such as castor oil and metallic soaps, had failed to give proper protection to the hypoid gear surfaces of hard ferrous alloy, it was a short but important step to the introduction by Evans (1937) of oil-soluble organic sulphur compounds which form surface films capable of preventing metal-to-metal contact under severe conditions of both loading and sliding speeds. Since then a bewildering variety of so-called 'extreme-pressure' additives has been increased by compounds containing chlorine or phosphorus. In 1942, Evans showed that the additive molecules did not perform their function by merely becoming adsorbed as such on the bearing surface, but found that a useful working hypothesis was to suppose that sulphide, chloride or phosphide layers were formed. Since then the formation of an unidentified iron sulphide compound, possibly FeS_2

(Simard, Russell and Nelson 1941), of an iron phosphide (Fe_3P) (Beeck, Givens and Williams 1941, Brummage 1942), ferrous chloride ($FeCl_2$) (Brummage 1942) and ferric oxychloride (FeOCl) (Brummage 1949, private communication) on ferrous bearing surfaces lubricated by oils containing the appropriate additives has been observed. Other layers produced in a similar manner have been identified on α-brass (cuprous chloride, CuCl) (Bryant 1936) and on silver (silver sulphide, Ag_2S). The formation of a sulphur compound layer has also been confirmed by the use of an additive containing radioactive sulphur (Clark, Gallo and Lincoln 1943). Thus Evans' view that the extreme-pressure additive forms a layer by a chemical attack can be accepted. Up to the present, however, no reasonable hypothesis has been put forward to account for the remarkable resistance of this layer to rupture. Indeed, wear of the properly lubricated hypoid gear is so slight that metal-to-metal contact and welding must be of rare occurrence.

Now, it is well known that an injury to the oxide layer on, for example, a ferrous bearing surface is almost immediately healed on renewal of the boundary layer owing to the action on the exposed metal of air and other oxidizing products dissolved in the ordinary lubricant. It is clear, therefore, that the superior protection afforded by the sulphide (or chloride or phosphide) layer formed by the extreme-pressure additive is due to some property inherent in the nature of this surface compound. It cannot be a question of hardness because, excepting pyrite, the sulphides, chlorides and phosphides of iron are softer than the oxides. A significant fact, however, is that, with few exceptions (the noble and platinum-group metals, tantalum, molybdenum and tungsten), the oxides likely to be formed on bearing surfaces are more refractory than the corresponding metals. On the other hand, the sulphides, chlorides and phosphides of the chief constituents of the alloy steels used in hypoid gears are more, or even much more, fusible than the metals. Finally, a zone of sliding contact between the oxide layers is a region of high temperature and, possibly, high pressure.

With these facts in mind, it seems to me that the extreme-pressure additive acts in the following way. Before the frictional temperature at areas of rubbing contact becomes high enough to soften the underlying metal and thus rob the oxide layer of its support, the additive reacts with the oxide to form a layer of a compound, such as a sulphide or oxysulphide, which is more fusible than the metal. If the local temperature should rise sufficiently to melt the compound, some of the molten layer will smear and thus act for a time as a lubricant. The one or two molecular layers next to the metal, being more strongly held, will, however, remain intact and continue to resist rupture, because the melting point of the compound is too low to impair the rigidity of the metal and thus deprive these last molecular layers of their support. This view is further supported by Greenhill's (1948) experiments which showed that the friction between sliding pairs of steel or copper surfaces was much reduced on replacing the oxide layers by sulphides.

It has already been remarked that it is a common experience that in time both bearing surfaces of a sliding element become polished, no matter how widely their melting points or hardnesses may differ. At first sight this seems to conflict with the results obtained by Bowden and Hughes (1937) who found that a soft material could polish a hard one of lower melting point without itself being polished.

The explanation lies in the difference between the conditions under which their experiments were carried out and those of normal sliding-surface practice. There are, indeed, conditions in which hardness and fusibility are of secondary importance. For example, platinum (H.4 : M.P.1,775°) is polished by a high-speed buff dressed with rouge (H.6½ : M.P. 1,565°). A diamond tool tip (H.10 : M.P. >3,500°) is worn and smoothed by corundum (H.9 : M.P. 2,050°) during the turning of aluminium. A rotating pad dressed with corundum polishes periclase (H. 5½–6 : M.P. 2,500–2,800°). Synthetic fibres wear and polish the hard steel surfaces over which they slide in textile machinery. The end of a rod of camphor (M.P. 178° c.) is caused to flow and be polished by a swiftly moving ribbon of Wood's metal (M.P. 73° c.) (Spurr, recent experiments carried out in this laboratory). These examples suffice to show the importance of the effect of a continual changing or otherwise of the rubbing surface. When two surfaces are in rubbing contact, the heat developed is shared by both. If one surface is always in rubbing contact, e.g. the tool tip, its temperature will exceed that of the other surface which is continually changing, despite conduction losses. For example, a small steel specimen describing a continually changing path on a brass surface can be run without seizure under much more severe conditions than vice versa (Brummage, private communication). In both cases, the small specimen reaches a high temperature, which the steel, because of its higher melting point, is better able to withstand. This is precisely what happens on running in a sliding element pair consisting of a hard and a soft metal. The soft metal is first polished and smoothed. The asperities on the hard surface then form small surfaces which continue to rub in relatively long contact paths over rounded hill tops on the softer surface and are thus in their turn smoothed and polished. In general, where high spots are in sliding contact, the one of smaller radius of curvature will attain the highest temperature, and this may well exceed the melting point of the layer on the surface of larger radius of curvature.

§ 6. THE BOUNDARY LAYER

The function of the oxide layers is to prevent intermetallic friction with its more or less destructive consequences. The oxide-layer (or sulphide etc.) component of friction, however, is still too high to be tolerated in general sliding practice, even supposing the layers could remain intact. It is the function of the boundary layers of the lubricant to replace the oxide-layer component of friction by one of still lower value.

The lubricants consist, in general, of mixtures of straight-chain hydrocarbon molecules (paraffins) or their end-substituted products (alcohols, esters, ketones, acids). At temperatures below the melting point of the hydrocarbons, the chains, 4·5 A. apart, are grouped in bundles, like sticks in fascines. Such bundles are in fact crystals. When grown from the melt or solution on a solid surface the crystals are orientated with the chains either vertical or inclined to the surface at an angle characteristic of the compound. The thickness of such an orientated crystalline layer may be as much as 50 molecules or more (for references see Brummage 1947).

When such layers are heated, as they are under practical conditions of sliding, the fascines break up at a temperautre close to the melting point, leaving a true boundary layer of molecules. The molecules attached directly to the surface are normal or steeply inclined to it; above them is a layer of less pronounced orientation, and by the time the fourth layer is reached the orientation is practically non-existent (Finch and Zahoorbux 1937); that is to say, the film has now become fluid. The maximum thickness of the boundary layer is therefore of the order of less than 200 A.

The chains of the first layer in the boundary film are all adsorbed with one end attached to the oxide film, the radicle end in the case of end-substituted molecules. Owing to the polar nature of these, they are more firmly held than the normal paraffins. Further, acids or esters may react with the oxide, sulphide or other layer to form soaps; in this case the relatively thick crystalline layer persists up to the softening temperature of the soap (Brummage 1947, Tingle 1950).

The transition from a crystalline to a boundary film is reversible in that, on cooling, the crystals are re-formed (Brummage 1947). In consequence, new orientations of the crystals or fascines are observed after rubbing. For example, solidified fatty acid films are frequently found to contain crystals in which the molecules are aligned in the direction of sliding at an angle of about 5° to the surface (Finch 1938, Germer and Storks 1939).

Boundary-layer friction is lower with end-substituted than with normal paraffins of similar chain lengths (Hardy 1936). This has generally been attributed to their stronger adhesion to the oxide layer. It is difficult, however, to see how this can materially reduce friction. It may be said, indeed, that present views on the mechanism of the facilitating of sliding by the lubricant boundary layer are inadequate. It is true that I have at times likened the boundary layer to a velvet pile or the bristles of a brush, and have also drawn attention to the thermal motion of the molecules of the layer and to possible effects of their polarity. But all these ideas are at best working hypotheses meant only to serve as guides to further experimental enquiry.

§7. THE INTER-BOUNDARY LUBRICANT FILM

The inter-boundary lubricant film exists only in regions of relatively light loading. Osborne Reynolds (1886) recognized its fluid character, and his theory of hydrodynamic flow not only represents the first broad generalization in the science of lubrication but has also dominated the geometrical design of the sliding element to this day. This is as it should be, for the shear strength of the fluid film is less than that between any of the other surface pairs involved in the sliding element. Boundary conditions of sliding should therefore be restricted as far as possible to those areas of heavier loading where the fluid film can no longer survive as such.

In addition to keeping the boundary layers apart and reducing friction in the more lightly stressed regions, the fluid film has other important functions to fulfil. It must replenish supplies to the boundary layers, bring up oxidizing agents (oxygen, sulphur, halogens etc.) to repair damage sustained by the oxide layers or suitably to modify these, and carry off debris, clothing each particle with a protective boundary layer. Finally, the fluid layer also helps to carry off frictional heat.

The main sources of wear in hydrodynamic lubrication are chemical corrosion and erosion. Erosive wear is chiefly due to debris rolling (Figure 6) between or striking sliding surfaces, and to cavitation (Munroe; see Rayleigh 1917) effects resulting from the collapse of bubbles suspended in the oil and in contact with the sliding surface.

§ 8. CONCLUSION

The problem of the sliding element is a complex one. A distinguished physicist once spoke of it to me as 'a very messy business'. This complexity together with the success of the empirical approach has undoubtedly deterred scientific research into the subject. Since Osborne Reynolds' triumphant framing of the theory of hydrodynamic lubrication, only four schools have attempted to carry the matter further. Beilby's (1921) work on surface flow and Hardy's (1936) views on surface forces were destined, as we have seen, to prove of fundamental importance. The other two schools, both active to-day, are those of Bowden in Cambridge and my own in South Kensington. Hitherto all these workers have been concerned mainly with the metallic bearing surface and the oil *qua* lubricant. If, in this lecture, I have laid particular emphasis on the oxide layer and on the oil as purveyor of oxidant to the high temperature zone of solid-to-solid sliding contact, it is because I believe that it is in this direction that the next great advance will be made. Without the oxide layer and oil-borne oxidants, whose interaction is confined to regions of high temperature, sliding under boundary conditions would be a practical impossibility.

ACKNOWLEDGMENTS

I have to thank Professor E. N. da C. Andrade and Dr. J. W. Fox for Figure 20, M. M. Bluhm for Figure 6, Dr. K. Leu for Figures 16 and 17, D. P. D. Webb for Figure 2, Dr. H. Wilman and D. Layton for Figure 3 and R. T. Spurr for Figures 5, 8 and 9 and for his help, particularly with the single crystal experiments.

My thanks are also due to Dr. K. G. Brummage (Anglo-Iranian Oil Company's Research Laboratories) and to Mr. E. A. Evans (C. C. Wakefield Company's Research Laboratories) for helpful discussions, to Mr. D. Green (Vandervell Limited) for the loan of bearing liners exhibited at my lecture and to Mr. G. Whipple (Hilger and Watts Ltd.) for the loan of a set of slip gauges.

I also wish to record my thanks for financial assistance given to the work carried out in this field in my laboratories by E. G. Acheson Ltd., Anglo-Iranian Oil Company Ltd., Ferranti Ltd., C. C. Wakefield Ltd., the Department of Scientific and Industrial Research, and the University of London Central Research Fund (for the provision of a 200 kv. generator).

REFERENCES

Amontons, G., 1699, see *Encyclopaedia Britannica*, 13th Ed., 1926.

Andrade, E. N. da C., and Fox, J. W., 1949, *Proc. Phys. Soc.* B, **62,** 483.

Beeck, O., Givens, J. W., and Williams, E. C., 1941, *Proc. Roy. Soc.* A, **177,** 103.

Beilby, G., 1903, *Proc. Roy. Soc.* A, **72,** 218; 1921, *Aggregation and Flow of Solids* (London: Macmillan).

Bowden, F. P., and Hughes, T. P., 1937, *Proc. Roy. Soc.* A, **160,** 575.

Bowden, F. P., and Ridler, K. E. W., 1936, *Proc. Roy. Soc.* A, **154,** 640.

Bowden, F. P., Stone, M. A., and Tudor, G. R., 1947, *Proc. Roy. Soc.* A, **188,** 329.

Bowden, F. P., and Tabor, D., 1939, *Proc. Roy. Soc.* A, **169,** 391; 1945, *Annual Reports Chem. Soc.*, **42,** 20.

Brummage, K. G., 1942, *Ph.D. Thesis, London University*; 1947, *Proc. Roy. Soc.* A, **188,** 414, **191,** 243.

Bryant, F. J., 1936, *Ph.D. Thesis, London University.*

Cabrera, N., and Mott, N. F., 1949, *Rep. Prog. Phys.*, **12,** 163.

Clark, G. L., Gallo, S. G., Lincoln, B. H., 1943, *J. Appl. Phys.*, **14,** 428.

Cochrane, W., 1938, *Proc. Roy. Soc.* A, **166,** 228.

COULOMB, C. A., 1781, see *Encyclopaedia Brittanica*, 13th Ed., 1926, article on ' Friction ' by A. B. W. Kennedy and W. E. Dalby.

DE BROUCKÈRE, L., 1945, *J. Inst. Metals*, **71**, 131.

DE LA RIVE, A. A., 1839, *Ann. Chim.* (*Phys*)., **39**, 328.

EVANS, E. A., 1937, *Proc. Instn. Mech. Engrs.*, General Discussion on Lubrication; 1942, *Prcc. Instn. Auto. Engrs.*, **37**, 1.

EWING, A., 1912, *J. Inst. Metals*, **8**, 30.

FARADAY, M., 1838, see *Experimental Researches in Chemistry and Physics* (London : Taylor and Francis, 1859).

FINCH, G. I., 1937, *Science Progress*, **31**, 609; 1938, *J. Chem. Soc.*, 1144; 1945, *Proc. Instn. Mech. Engrs.*, **153**, 331.

FINCH, G I., and QUARRELL, A. G., 1933, *Proc. Roy. Soc.* A, **141**, 398.

FINCH, G. I., QUARRELL, A. G., and ROEBUCK, J. S., 1934, *P.oc. Roy. Soc.* A, **145**, 676.

FINCH, G. I., QUARRELL, A. G., and WILMAN, H., 1935, *Trans. Faraday Soc.*, **31**, 28.

FINCH, G. I., and SUN, C. H., 1936, *Trans. Faraday Soc.*, **32**, 852.

FINCH, G. I., and WHITMORE, E. J., 1938, *Engineering*, **146**, 91.

FINCH, G. I., and WILLIAMS, A. L., 1937, *Trans. Faraday Soc.*, **32**, 852.

FINCH, G. I., and WILMAN, H., 1936, *Proc. Roy. Soc.* A, **155**, 345.

FINCH, G. I., WILMAN, H., and YANG, L., 1947, *Discussions, Faraday Soc.*, **43A**, 144.

FINCH, G. I., and ZAHOORBUX, F. D., 1937, *Proc. Instn. Mech. Engrs.*, General Discussion on Lubrication.

GERMER, L. H., and STORKS, K. H., 1939, *Phys. Rev.*, **55**, 648.

GREENHILL, E. B., 1948, *J. Inst. Petrol.*, **34**, 659.

HARDY, W. B., 1936, *Collected Scientific Papers* (Cambridge : University Press).

HARDY, W. B., and HARDY, J. K., 1919, *Phil. Mag.*, **38**, 32.

HOPKINS, B. F., and PALM, J. V. O., 1932, British Patent No. 409,289 (accepted 1934).

HOPKINS, H. G., 1935, *Trans. Faraday Soc.*, **31**, 1095.

LANGMUIR, I., 1916, *J. Amer. Chem. Soc.*, **38**, 2269; see also, 1920, *Gen. Elect. Rev.*, **23**, 503.

PARKER, R. C., and HATCH, D., 1950, *Proc. Phys. Soc.* B, **63**, 185.

PRESTON, G. D., and BIRCUMSHAW, L. L., 1936, *Phil. Mag.*, **22**, 654.

RAYLEIGH, Lord, 1917, *Phil. Mag.*, **34**, 94.

REYNOLDS, O., 1886, *Philos. Trans.*, **177**, 156.

SIMARD, G. L., RUSSELL, H. W., and NELSON, H. R., 1941, *Industr. Engng. Chem.*, **33**, 1352.

THIRSK, H. R., and WHITMORE, E. J., 1940, *Trans. Faraday Soc.*, **36**, 565, 862.

TINGLE, E. D., 1940, *Trans. Faraday Soc.*, **46**, 93.

VERNON, W. H. J., WORMWELL, F., and NURSE, T. J., 1944, *J. Iron and Steel Inst.*, **150**, 81.

WILLIAMS, C. G., 1937, *Proc. Instn. Mech. Engrs.*, Discussion on Lubrication.

WILMAN, H., 1950, *Nature, Lond.*, **165**, 321, *Proc. Phys. Soc.* A, **63**, 298.

The Influence of Concentration-Dependent Diffusion on Rate of Evaporation

By J. CRANK
Courtaulds Limited, Maidenhead, Berks.

MS. received 18th January 1950

ABSTRACT. It has been suggested that evaporation through a membrane or from a polymer film has a characteristic behaviour different from that commonly associated with a free liquid surface. In particular it is claimed that an increase in surface humidity produces an increased rate of evaporation through a membrane, though the opposite effect is well known for a liquid surface. This behaviour is said to be observed when diffusion within the membrane or film is the rate controlling process in systems for which the diffusion coefficient increases with increasing concentration.

In this paper the behaviour of such diffusion-controlled evaporation processes is examined from the point of view of the diffusion equations and it is found that there is no theoretical reason to expect such processes to show anomalous behaviour.

It is seen that when the experimental data for evaporation through a membrane are correctly interpreted they are in agreement with theoretical prediction.

The so-called ' skin effect ' in the dry spinning of fibres is examined and appears to be due to complicating features other than diffusion.

§ 1. INTRODUCTION

Many evaporative processes are diffusion-controlled in the sense that the rate of evaporation depends largely on the rate at which solvent or moisture is supplied to the evaporating surface by internal diffusion. The dry spinning of fibres and the loss of moisture from the body through the skin are examples of particular interest to the textile industry.

It is of considerable practical importance to know how the rate of evaporation is influenced by external conditions, particularly those over which control can be exercised. Thus in the evaporative (dry) spinning process, for example, not only the rate of production of fibres but also the shape of the filaments and their properties are profoundly affected by the rate of evaporation of solvent at different stages of the process.

The present paper is concerned with the effect of the proportion of solvent vapour, or the relative humidity, in the atmosphere into which the evaporation takes place. In the case of evaporation from a free liquid surface it is well known that the higher the vapour pressure in the external atmosphere the less the rate of evaporation and vice versa. Correspondingly, the higher the wind speed past the exposed surface, the greater the rate of evaporation because the thickness of the boundary layer through which the vapour must pass in evaporating decreases with increasing wind speed. With this in mind it is somewhat surprising to find that several instances of the opposite behaviour have been reported, where an increase in relative humidity at the evaporating surface produces an increased rate of evaporation. This is said to happen when evaporation occurs through an organic membrane or from a polymer film, in which cases the supply of solvent to the surface is governed by diffusion. The following are examples of this anomalous behaviour.

Throughout the patent literature of dry spinning (see for example British Patent Specification 238,842) there is the suggestion that the rate of loss of solvent from the filaments, e.g. of acetone from cellulose acetate filaments, can be increased by including a large proportion of solvent vapour in the atmosphere through which the filaments pass after issuing from the dies. This behaviour is attributed to the fact that when the evaporation takes place into an atmosphere containing little or no solvent vapour a dry surface layer, frequently called a skin, is formed almost as soon as the filaments leave the dies. As it is known that the diffusion coefficient is very small in a near-dry polymer, the skin is considered to act as a barrier to the subsequent diffusion of solvent from the filament core and hence to slow down the evaporation in the later stages. The formation of the skin is prevented by using a drying atmosphere rich in solvent vapour, and it is therefore argued that a more rapid removal of solvent from the filaments is obtained.

More recently Cassie (1949), discussing the characteristics for warmth in underwear fabrics, has stated that, at low temperatures, the heat loss by evaporation from the body increases as the relative humidity surrounding the skin increases. This he considers to be a well-known effect for membranes such as the skin and in support of the observation refers to the diffusion coefficient for water through a keratin membrane, the coefficient being relatively very small at low moisture contents.

The purpose of the present paper is to show mathematically that, if such anomalous behaviour exists, it cannot be attributed simply to the low value of the diffusion coefficient in the near-dry film or membrane. Calculations show that any abnormality in behaviour must be due to complicating features other than the simple diffusion process. In order to reduce the mathematical complications to a minimum, the main calculations are carried out for a lamina instead of for a cylinder. Effects such as puckering of the surface of the filaments are neglected.

Part I of the paper deals with the steady-state flow through a membrane, and Part II with the rate of the loss of solvent from a polymer film.

PART I

§ 2. STEADY-STATE EVAPORATION THROUGH A MEMBRANE

(i) *Theoretical*

If we have a membrane of thickness l separating a region of high relative humidity from one of low relative humidity, then, provided the humidities on the two sides are maintained constant, a steady-state transfer of moisture through the membrane is set up. The concentrations C_0 and C_1 just within the two surfaces of the membrane are constant and are the equilibrium regains for the high and low humidities respectively. A concentration gradient exists through the membrane, and moisture moves down the gradient by diffusion at a rate determined by the concentrations C_0 and C_1 and the diffusion coefficient D. The rate of evaporation F per unit area of the membrane is given by

$$F = -D\,dC/dx, \qquad \text{......(1)}$$

which on integration becomes

$$\int_0^l F\,dx = \int_{C_1}^{C_0} D\,dC. \qquad \text{......(2)}$$

Since we are concerned with a steady state, F is the same through each section of the membrane, i.e. F is independent of x. Therefore we have

$$\int_0^l F\,dx = F\int_0^l dx = Fl, \qquad \ldots\ldots(3)$$

and finally

$$F = \frac{1}{l}\int_{C_1}^{C_0} D\,dC. \qquad \ldots\ldots(4)$$

The above argument is true whether D is constant or not. In particular it is true when D is a function of concentration, and provided D is never negative in the range from C_0 to C_1, the integral in (4) must always increase or remain constant as the range of concentration is increased. Thus if the high concentration C_0 remains fixed, the rate of flow F can never decrease as a result of decreasing the lower concentration C_1. The significance of (4) in this connection was brought to my notice by my colleague, Dr. G. S. Park.

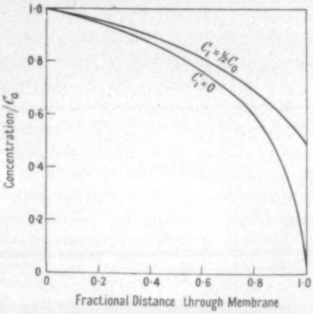


Figure 1. Concentration distribution through a membrane for a diffusion coefficient increasing exponentially with concentration. C_0, C_1 are the concentrations on the high and low pressure sides respectively.

Figure 1 shows calculated concentration distributions through the membrane when D is an exponential function of C such that D increases by 50-fold from $C=0$ to $C=C_0$. The concentration C_0 on the high humidity side is the same for each curve but in one case $C_1=0$ and in the other $C_1=\frac{1}{2}C_0$. The curves show how the concentration distribution adjusts itself so that the rate of flow can be greater when $C_0=0$ even though the region of low diffusion coefficient is then included. Clearly a low diffusion coefficient is compensated by a high concentration gradient.

(ii) *Experimental*

In a recent paper Cassie (1949) remarks that it is well known that an increase in relative humidity on the low humidity side of a membrane results in an increased rate of evaporation. He describes a convenient way of demonstrating the effect experimentally, by observing the rate of evaporation from a collodion bag filled with water and reports that "the evaporative loss from the bag actually decreases as the wind speed increases because the higher wind speed maintains a low relative humidity at the surface of the bag". The behaviour described appears plausible on the grounds of the intuitive reasoning outlined earlier in the paper, i.e. that increasing the surface humidity prevents the formation of a near-dry surface layer through which diffusion is slow. Thus there appears to be a discrepancy between the behaviour predicted by equation (4) on the one hand, and that expected intuitively and observed experimentally on the other.

The original observations were made by using the collodion bag as a wet bulb thermometer. An increased wind speed was accompanied by an increase in temperature and from this it was concluded that the rate of evaporation had decreased. The experiment has recently been repeated, however (Cassie, private communication), and, whilst the temperature change is confirmed, a direct measurement of the rate of evaporation reveals that this does in fact increase as the wind speed increases. This is in agreement with the mathematical prediction of equation (4) above. If a temperature increase accompanies the increase of wind speed (probably due to an increase in heat exchange from the moving air) it follows from (4) that the observed increase in evaporation will be greater than if the temperature remained constant, because the diffusion coefficient D increases with increasing temperature. Thus we conclude that the rates of evaporation through a membrane and from a free liquid surface are both decreased by increasing the surface humidity or decreasing the wind speed.

PART II

§3. EVAPORATION OF SOLVENT FROM POLYMER FILMS AND FILAMENTS

Physical Model on which Calculations are based

It is assumed that the loss of solvent from the polymer film takes place by diffusion through the film with subsequent evaporation from the surface exposed to the drying air stream. On this assumption the course of drying may be calculated from the diffusion equation. We have seen that the diffusion coefficient is generally concentration dependent in such systems, and the nature of this variation and also the law governing the surface evaporation must be known before the calculations can proceed.

The calculated results shown below relate to the loss of acetone from a plane sheet of cellulose acetate exposed to a drying atmosphere on one surface only. The concentration of acetone is initially uniform throughout the sheet and constant temperature conditions are assumed. Two kinds of condition at the exposed surface are considered:

(i) The surface of the sheet is assumed to reach equilibrium with the drying atmosphere instantaneously when drying commences, i.e. if the atmosphere is acetone-free, the concentration at the surface falls instantaneously to zero; if the vapour pressure in the atmosphere here is p, the surface of the sheet immediately reaches the concentration which is in equilibrium with p.

(ii) A condition expressing the rate of evaporation of acetone from the surface is assumed. The simplest reasonable condition is that the rate of evaporation is directly proportional to the difference between the actual concentration, C_s, in the surface of the sheet at any time and the concentration C_0, which would be in equilibrium with the vapour pressure in the atmosphere remote from the surface. Mathematically this means that at the surface

$$D_s \partial C/\partial x = \alpha(C_s - C_0), \qquad \ldots\ldots(5)$$

where α is the constant of proportionality and D_s is the value of the diffusion coefficient when the concentration is C_s; x is the space coordinate perpendicular to the face of the sheet, increasing positively inwards.

The calculated rates of loss of solvent given below are obtained by evaluating solutions of the diffusion equation

$$\frac{\partial C}{\partial t} = \frac{\partial}{\partial x}\left(D\frac{\partial C}{\partial x}\right), \qquad \ldots\ldots(6)$$

with a surface condition of the type (i) or (ii) above. Only the results of the calculations are given here. Details of methods of evaluating numerical solutions of the diffusion equation with a variable diffusion coefficient, and taking account of any shrinkage of the film accompanying loss of solvent, are described elsewhere (Crank and Henry 1949).

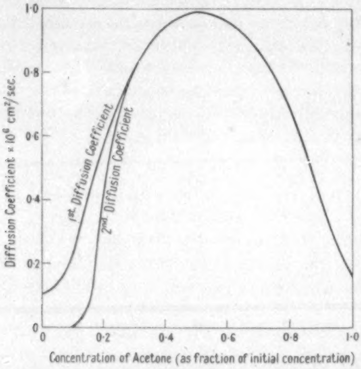


Figure 2.

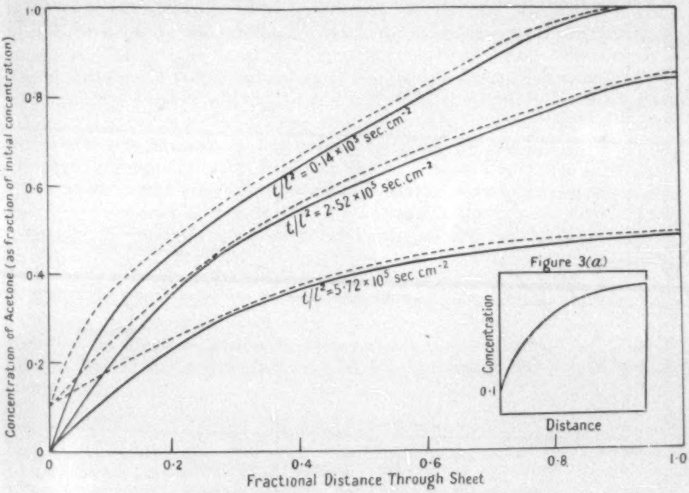
The first diffusion coefficient plotted in Figure 2 is derived from observations of the rate of penetration of acetone into commercial cellulose acetate sheet, using an interferometer technique described by Crank and Robinson (1950). The significance of the second diffusion coefficient shown in Figure 2 is discussed later.

§ 4. CALCULATED RATES OF DRYING AND DISTRIBUTION OF ACETONE

(i) *Surface taking up its final concentration instantaneously.*

This is the condition of vigorous surface evaporation. Figure 3 shows the calculated variation of concentration of acetone with distance through the sheet at three successive times for two different surface concentrations; (1) the surface concentration falls to zero instantaneously, (2) it falls to one-tenth of its initial value instantaneously. There is little difference between the two curves at any time. In particular the total acetone content of the sheet at any time, represented by the area under the appropriate curve of Figure 3, is much the same whether the surface concentration falls to zero or to one-tenth its initial value, i.e. whether a concentration range over which D is small is removed or not. In so far as there is a slight difference, the drier sheet at any time is obtained by putting the surface to zero concentration, which means vigorous evaporation into an acetone-free atmosphere. No advantage is gained by introducing acetone into the atmosphere.

The reason for this is clear from the calculations, as it was in the steady-state case shown previously in Figure 1. When the value of D at the surface is very low the concentration gradient is correspondingly large so that the overall rate of loss of acetone, given by the product $D\partial C/\partial x$ at the surface, does not alter appreciably. This compensating effect is most obvious when the diffusion coefficient is zero over a range of low concentrations as in the second diffusion coefficient shown in Figure 2. This is a hypothetical coefficient chosen to exaggerate the effect. The concentration–distance curve for this coefficient is sketched in Figure 3 (*a*) from general reasoning. The surface gradient is infinite over the concentration range for which D is zero (in this case for concentrations less than 0·10) so that the product can have a non-zero value. A finite gradient cannot develop for concentrations less than 0·10, since this would imply that solvent had been removed from a region of zero diffusion coefficient under the action of a finite gradient, which is not possible.



——— Surface concentration zero.
– – – – – Surface concentration one-tenth of uniform initial concentration through sheet.

Figure 3. Concentration distributions for the first diffusion coefficient of Figure 2.

Thus if we interpret the term 'skin' to mean a gel whose acetone content is less than a certain critical amount, so that the diffusion coefficient in the skin is low, we find that the lower the mean diffusion coefficient in the skin the greater the concentration gradient throughout the skin which is therefore thinner. With this kind of surface condition, therefore, it appears that it is not possible to form, by diffusion of acetone, a skin which can seriously impede further diffusion of acetone.

(ii) *A rate of surface evaporation proportional to* $C_s - C_0$.

Figure 4 shows the concentration–distance curves calculated for the surface rate of loss condition (5) and the first diffusion coefficient of Figure 2. The curves of Figure 4 are for $\alpha l = 16 \times 10^{-6}$; in one case $C_0 = 0$ and in the other $C_0 = 0{\cdot}2$, measured as a fraction of the initial concentration in the sheet. At any given time the surface is always drier for $C_0 = 0$, but there is the rather interesting feature that the concentration–distance curves cross within the sheet. Thus there is an interval of time during which the inside of the film is drier at a given time when the evaporating atmosphere contains acetone than when it is acetone-free. The corresponding curves for $\alpha l = 8 \times 10^{-6}$ were calculated, and found not to intersect. They are not sufficiently interesting to show. Since the curves for vigorous evaporative conditions, i.e. $\alpha = \infty$, do not cross either, it is clear that there is a limited range of values of αl for which crossing is observed. Although the magnitude of this effect will vary slightly with different values of αl and C_0, and different diffusion coefficients, other calculations, detailed results of which are not shown, suggest that the effect will never be very large. Thus for the second diffusion coefficient in Figure 2 the separation between the concentration–distance curve is not appreciably different from that of the corresponding curves in Figure 4.

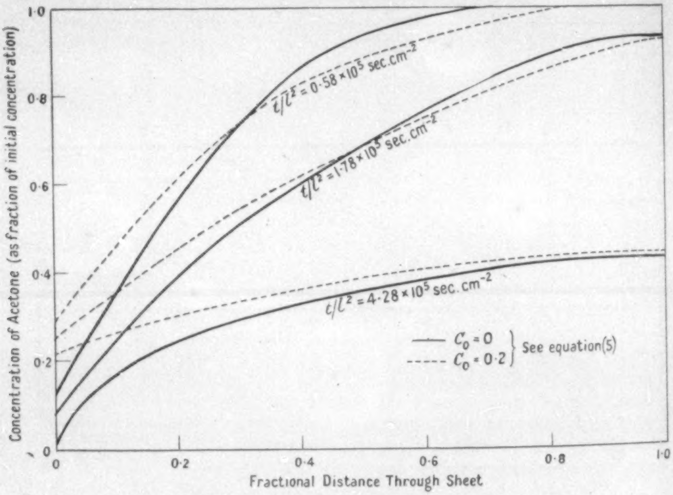


Figure 4. Concentration distributions for first diffusion coefficient of Figure 2, $\alpha l = 16 \times 10^{-6}$.

It is clear from Figure 4, as from Figure 3, that the overall rate of loss of solvent from the film is not significantly increased by introducing acetone into the atmosphere.

§ 5. CYLINDRICAL FILAMENTS

Similar calculations for the rate of loss of solvent from a cylinder show the same general behaviour as for the plane sheet. The most rapid overall loss of solvent is obtained when the drying atmosphere is solvent-free. The crossing of the concentration–distance curves is even less marked than in Figure 4.

§ 6. CONCLUSION

Theoretical considerations of diffusion-controlled evaporation do not appear to support the view that the rate of evaporation either through a membrane or from a polymer film can be increased by increasing the surface humidity. In the case of the membrane, the discrepancy between theory and experiment seems to have been due to a misinterpretation of the observational data and this discrepancy has now been removed.

If the presence of solvent vapour in the atmosphere in other cases really does result in an increased rate of evaporation as reported, it appears that this must be due to factors other than simple diffusion. In the case of the dry spinning process, for example, the tension to which the filaments are usually subjected whilst losing solvent may be an important complicating feature. The degree of stretch, and hence of radial contraction, produced by a given tension must depend very much on whether a near-dry skin is formed or not in the early stages of drying. Thus the main effect of solvent vapour in the atmosphere may be to allow the filament cross section to be decreased considerably by longitudinal stretching. The rate of drying is thereby increased since it is inversely proportional to the square of the radius. However, a detailed investigation of the cause of any apparently anomalous behaviour is outside the scope of the present paper. In all probability the explanation will differ from one system to another.

ACKNOWLEDGMENTS

It is a pleasure to thank Dr. A. B. D. Cassie for a most helpful correspondence concerning the observations on evaporation through membranes, and for his willingness to allow the essential points to be communicated in Part I of the paper. The author is indebted to Miss M. E. Henry who carried out the calculations.

REFERENCES

British Patent Specification No. 238,842, Société pour la fabrication de la soie rhodiasata.
Cassie, A. B. D., 1949, *J. Text. Inst.*, **40**, 444.
Crank, J., and Henry, M. E., 1949, *Trans. Faraday Soc.*, **45**, 636.
Crank, J., and Robinson, C., 1950, in course of preparation.

The E.M.F.–Temperature Calibration of Platinum, 10% Rhodium–Platinum and Platinum, 13% Rhodium–Platinum Thermocouples over the Temperature Range 0°–1,760° c.

By C. R. BARBER
Physics Division, National Physical Laboratory, Teddington, Middlesex

Communicated by Ezer Griffiths; MS. received 31*st October* 1949, *and in amended form* 16*th January* 1950

ABSTRACT. E.M.F.–temperature reference tables are given for platinum, 10% rhodium–platinum and platinum, 13% rhodium–platinum thermocouples over the range from 0° to 1,760° c. The tables are based on the calibration of twelve thermocouples of each kind, six from each of two manufacturers. Each set of six thermocouples comprised two samples from each of three batches of wire. The calibrations are in terms of the International Temperature Scale as revised at the Ninth General Conference of Weights and Measures held in Paris in October 1948.

The methods of intercomparison between the thermocouples and the various interpolation instruments specified in the International Temperature Scale are described. It is shown that the difference between the new tables and the old are mainly due to changes in the International Temperature Scale, and that the new tables represent the shape of the E.M.F.–temperature characteristics of the thermocouples of either make sufficiently closely to give interpolation between calibration points with accuracy within ±1° c. up to 1,063° c., ±2° c. up to 1,550° c., and ±3° c. above this temperature.

§ 1. INTRODUCTION

Reference tables for platinum, rhodium–platinum thermocouples of British origin have hitherto been prepared by the manufacturers of the thermocouples from skeleton tables given by the National Physical Laboratory in reports on the routine calibrations of sample thermocouples over a number of years, and from a reference table for 13% couples published in 1943 (Barber 1943) for the temperature range from 1,400° to 1,750° c. This range of temperature is of importance in steel melting, and the reference table was prepared to meet the needs of the then recently developed quick-immersion technique of liquid steel temperature measurement. Since the International Temperature Scale above 1,063° c. is realized by use of the optical pyrometer, the accurate determination of the form of the thermocouple calibration curve involves a comparison between thermocouple and optical pyrometer. In 1942 other urgent work made it impossible to do this, and less direct methods had to be relied upon.

A reference table of the kind here considered sets out to give the shape of the E.M.F.–temperature curve as precisely as possible. It may be used to interpolate accurately between a small number of calibration points on any particular thermocouple, in any desired temperature range, by plotting the difference of the calibration points from the values in the reference table. The shape of the E.M.F.–temperature curve varies very slightly from one thermocouple to another, presumably on account of minute traces of impurity in the thermocouple wires. The accuracy of the reference table in representing the curvature of the E.M.F.–temperature relation of any particular thermocouple is thus limited, and the greater these variations are the more calibration points are needed to compose a difference curve which will give an accurate calibration of an individual couple.

The tables given at the end of this paper have superseded those formerly issued by the makers, from which they differ quite markedly above the gold point (1,063° C.). This difference is mainly accounted for by the changes in numerical values of temperature in certain ranges resulting from the introduction, on 1st January 1949, of the International Temperature Scale (Comité internationale des Poids et Mesures, 1948), revising the original 1927 scale.

§ 2. SCOPE OF THE INVESTIGATION

The calibration of the thermocouples in terms of the International Temperature Scale involves a comparison ultimately with the interpolation instruments specified in the scale. For the range from 0° to 630° C. the platinum resistance thermometer is employed, using a quadratic equation,

$$R_t = R_0(1 + At + Bt^2),$$

relating resistance to temperature, based on the melting point of ice (0° C.), the boiling point of water (100° C.), and the boiling point of sulphur (444·60° C.).

From 630° to 1,063° C. the interpolation instrument is the platinum, 10% rhodium-platinum thermocouple, employing a quadratic equation,

$$e = a + bt + ct^2,$$

relating E.M.F. to temperature, based on the freezing points of antimony (630·5° C.), silver (960·8° C.) and gold (1,063° C.).

Above 1,063° C. the optical pyrometer is used and is calibrated on the basis of the freezing point of gold and the Planck law of radiation with a value of 1·438 for the constant C_2. A temperature t is then defined by the formula

$$\frac{J_t}{J_{\mathrm{Au}}} = \frac{\exp\{C_2/\lambda(T_0 + t)\} - 1}{\exp\{C_2/\lambda(T_0 + t_{\mathrm{Au}})\} - 1},$$

where J_t and J_{Au} are respectively the spectral energies radiated at wavelength λ by unit area of a black body in unit time at the temperature t and at the freezing point of gold, t_{Au}. T_0 is the melting point of ice in degrees K. defined for the purpose as 273·15.

The most significant differences in the scale outlined above compared with the 1927 scale are that the silver point is defined as 960·8° C. instead of 960·5° C., the value of C_2 is defined as 1·438 instead of 1·432, and the Planck Law of radiation is used instead of the Wien law. The effect of these differences on the numerical values of temperature is shown in Figure 1, where it is seen that in the region between 630° and 1,063° C. the temperatures as defined by the 1948 scale are higher than those of the 1927 scale, with a maximum difference of 0·4° C. at about 850° C., and above 1,063° C. they are lower, progressing from 0° at 1,063° C. to 1·6° at 1,400° C. and 3·6° at 1,700° C. These changes account for the larger part of the departures of the E.M.F. values given in the present reference tables from those formerly supplied by the makers. The adoption of the absolute electrical units in place of the international units on 1st January 1948 caused a rise in numerical values of voltage of 34 parts in 10^5, and this change must also be taken into consideration when the new tables are compared with the old.

§ 3. CALIBRATION OVER THE RANGE 0° TO 1,063° c.

The determination of the mean E.M.F.–temperature curves for both the 10% and 13% rhodium–platinum types of thermocouple has been based on the calibrations of twelve thermocouples of each kind, six being supplied by Messrs. Johnson, Matthey & Co. Ltd. and six by the Sheffield Smelting Co. Ltd. Each set of six couples comprised two from each of three batches of wire. They were annealed electrically at a temperature of about 1,600° c. and then calibrated with cold junctions in ice by direct comparison with platinum resistance thermometers over the range from 0° to 630° c. in stirred water, oil and salt baths (Grace and Hall 1943).

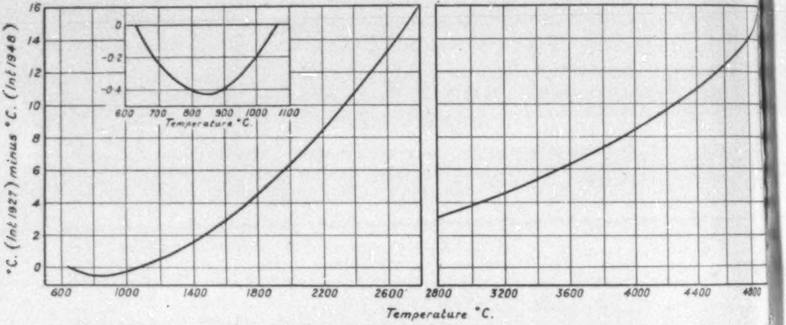


Figure 1. Changes in numerical value of temperature resulting from change from International Temperature Scale 1927 to International Temperature Scale 1948.

The measurements at the freezing point of silver were made on an ingot of pure silver of about 380 gm. weight, cylindrical in shape, of 2·5 cm. diameter and 10 cm. length. The ingot was contained in a graphite crucible and the surface of the ingot was covered with graphite powder to a depth of about 1 cm. The thermocouple was sheathed with a silica tube of 6 mm. external diameter and was immersed coaxially in the ingot to a depth of 9 cm., the silica sheath being protected from attack by the silver by a thin graphite sheath. Uniform temperature conditions over the length of the ingot were obtained by use of a substantial block of graphite for the crucible, which was formed by boring an axial hole 2·5 cm. in diameter in a cylinder about 7·7 cm. diameter and 20 cm. long. A series of freezes was observed on two of the thermocouples to establish the validity of the point. The reproducibility of E.M.F. of the thermocouples from one freeze to another was within 1 microvolt, equivalent to better than 0·1° c. An upward displacement of the thermocouple in the ingot of 2 cm. during a freeze produced no greater change than 0·2 microvolt. Having thus established that satisfactory readings were being obtained, a freezing point observation was made on each of the 24 thermocouples in turn.

For the gold-point observations, an ingot of 570 gm. weight, 2·5 cm. diameter and 6 cm. length was used. Since it is not necessary to protect the gold from the air as with silver, the graphite crucible was dispensed with and a silica crucible used with the thermocouple in a silica sheath immersed directly into the gold.

The crucible was placed in an axial hole bored in a nickel block in order to promote temperature uniformity. A similar degree of reproducibility was obtained as with the silver freeze.

§4. CALIBRATIONS FROM 1,063° TO 1,760° C.

The calibrations above the gold point involve comparison either directly or indirectly with the optical pyrometer. The indirect method is to use the freezing points of pure metals already determined on the optical pyrometer scale. The only points easily realized in the temperature range in question are the freezing points of palladium (1,552° c.) and platinum (1,769° c.). Other metals, such as nickel (freezing point 1,453° c.) and cobalt (freezing point 1,492° c.) must be melted *in vacuo* or in a reducing atmosphere. Even if all four of these points were used, however, a large temperature interval with no calibration point would be left between 1,063° c. and 1,453° c. The direct method was therefore adopted, and this required a black-body enclosure containing the hot junction of one or more thermocouples, the E.M.F. of which could be measured at the same time as the temperature of the enclosure was determined by means of the optical pyrometer.

A diagram of the furnace used is shown in Figure 2. The black-body enclosure was formed by a hollow platinum cylinder 2 cm. in diameter and wall thickness 1 mm. The back wall of the cylinder was set at an angle to the front wall so that the optical pyrometer should not receive a reflected image of the sighting aperture, which was a circular hole of 2 mm. diameter in the centre of the front wall. The furnace winding consisted of 10% rhodium–platinum foil 12·5 mm. wide and 0·025 mm. thick.

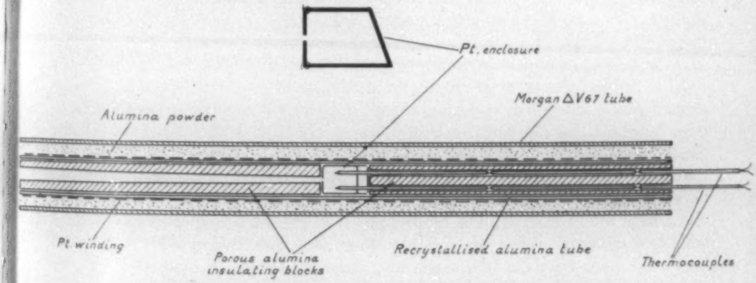


Figure 2. Central furnace tubes and black-body enclosure.

The optical pyrometer employed was the primary standard instrument in which the telescope is divided to permit the use of a rotating sectored disc between objective and lamp. It was housed in a constant temperature room and was calibrated immediately before these measurements at the fundamental point of the scale, the gold point, using the furnace arrangement shown in Figure 3. The furnace tube, 2 cm. in bore, was unlagged except for a single radiation shield. The gold ingot, weighing about 30 gm., was contained in a steatite crucible and had an alumina sighting tube of 2·5 mm. bore passing through the axis. In the centre of this tube was a steatite plug through which was drilled a very fine hole.

The furnace was used in the horizontal position and, since it was symmetrical, could be viewed from either end. The more usual form of black-body tube for this purpose is closed at one end and is immersed vertically in an ingot of the metal. It is uncertain with this method that the pyrometer is sighted along the axis of the tube since there is no visible detail in the furnace when uniform temperature conditions exist. The difficulty is overcome in the present arrangement, for the hole in the plug appears as a black spot in the field of the pyrometer when the alignment is correct. As there is only a small amount of furnace lagging, the end-cooling is small in comparison with radiation loss and, since the ingot is relatively short, very uniform temperature conditions are obtained. This is further aided by the fact that the thermal inertia of the furnace is small, and it is therefore only necessary to change the temperature of the windings by a few degrees in order to cause the metal to melt or freeze, so that the temperature gradients are limited. Under these conditions well-defined melting- and freezing-point curves were obtained and no significant differences were detected between observations made by sighting into opposite ends of the furnace.

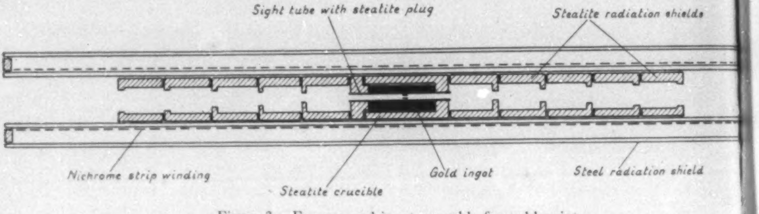


Figure 3. Furnace and ingot assembly for gold point.

The field of an optical pyrometer sighted into such a furnace is not uniformly bright. When the ingot is freezing the black-body tube appears as a bright central spot surrounded by darker zones and, when it is melting, as a dark spot with brighter surroundings. Under these conditions the matching of the filament against the central spot is influenced by the differences in intensity of the adjacent parts of the field, and in this determination led to an apparent difference between melts and freezes of 0·7° c. By using a supplementary lens to give a real image of the furnace, and at this position placing a diaphragm of slightly smaller diameter than the image of the black body, the field appears as a central bright spot in complete blackness when the pyrometer is focused on to the diaphragm; agreement between melts and freezes is then obtained.

The pyrometer was then sighted into the platinum enclosure, into which had been inserted thermocouples calibrated at the freezing point of gold as already described. When the furnace was adjusted to give the gold-point temperature as indicated by the E.M.F. of the thermocouples, the optical pyrometer readings agreed, to within 0·3° c., with those obtained by the direct method, thus giving a check on the efficiency of the black-body radiator. A further check was obtained in a similar manner at the freezing point of palladium (1,552° c.), the thermocouples having been calibrated at that point by the wire method and the pyrometer by use of the gold point and a sectored disc of appropriate aperture. The mean difference between the two methods in this case was only about 0·1° c.

Selected thermocouples were calibrated against the optical pyrometer in pairs at temperatures of about 1,200°, 1,270°, 1,310°, 1,420°, 1,550° and 1,650° c., each of these temperatures, with the exception of the last-mentioned, being determined by use of a sectored disc of appropriate aperture to cut down the spectral brightness of the black body, at the temperature named, to that at the gold point. The reduction from 1,650° c. was achieved in two stages, 1,650°–1,200° and 1,200°–1,063° c. The final measurement at the platinum point by the wire method, in which the platinum arm of the thermocouple is melted at the hot junction, was carried out in a manner similar to that previously described (Barber 1943). The couples calibrated in this way were compared with the remainder, after each point, in a separate furnace held at the same temperature. At the same time, as a check for deterioration in E.M.F., comparison was made with a control couple at the gold point and no significant changes were detected. On completion of all the observations up to and including that at 1,650° c., the E.M.F. of the thermocouples at the gold point had not changed by more than two microvolts from the original values.

§ 5. DISCUSSION OF RESULTS

Figure 4 shows the departures in E.M.F. of the mean of the six thermocouples of each make from the mean of all twelve thermocouples of one kind. These departures nowhere exceed 5 μv. for either the 10% or the 13% types, and common reference tables may therefore be computed without significant loss in accuracy in representing the true shapes of the E.M.F.-temperature curves. The departures in E.M.F. of individual couples of the 13% type of each make from their own means are shown in Figures 5 and 6. The 10% couples showed similar deviations. Since the wires were not specially selected, the deviations may be taken as representative of those likely to be obtained with any other couples from the same sources.

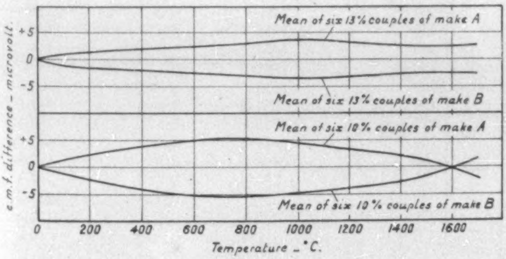


Figure 4. Departures of mean of E.M.F. of six couples of each make from mean of all twelve.

The means of the observations for the twelve 10% and for the twelve 13% couples are recorded in Table 1, and from these values the reference Tables 2 and 3 have been computed by the method described in the Appendix. The difference between the E.M.F. values given by the reference tables and the observational values are shown in columns 4 and 6 of Table 1. Apart from the values at 1,270° and 1,309° c. on the 10% couples and 1,270° c. on the 13% couples, there is no departure greater than 3 μv. (0·2° c.).

The accuracy of the measurements is estimated to be to within ±0·1° c. up to 1,063° c., ±0·2° c. at the gold point as measured with the optical pyrometer, rising to ±2° c. at 1,650° c. The calibration of a particular thermocouple obtained by measurement at a few temperatures and the use of a difference curve from the reference table will include these inaccuracies and any occurring in the calibration points. On the basis of these considerations it is estimated that observations at the gold and palladium points by the wire method and the use of the tables should yield an accuracy within ±1° c. up to 1,063° c., ±2° c. up to 1,550° c. and ±3° c. above.

Table 1

Mean Observed Values of E.M.F. of twelve 13% and twelve 10% Thermocouples

Comparison instrument	Temp. (° c.) (Int. 1948)	13% couples		10% couples	
		E.M.F. (μv.)	Diff.* (° c.)	E.M.F. (μv.)	Diff.* (° c.)
Platinum wire	1769·0	21107	0·0	18692	−0·2
Optical pyrometer	1648·5	19514	0·0	17334	0·0
	1552·0	18193	0·0	16204	−0·1
	1420·9	16339	−0·1	14622	+0·2
	1308·8	14745	−0·3	13270	−0·6
	1270·3	14192	+0·6	12787	+0·6
	1201·0	13239	−0·2	11958	0·0
	1129·0	12242	+0·2	11098	+0·1
Gold ingot	1063·0	11351·9	−0·1	10325·3	−0·1
Silver ingot	960·8	9990·9	−0·2	9137·9	0·0
Resistance thermometer	630·5	5919·5	0·0	5539·5	+0·1
	550·0	5009·7	0·0	4719·4	0·0
	500·0	4459·5	0·0	4219·8	0·0
	450·0	3920·7	0·0	3730·5	−0·1
	400·0	3397·1	0·0	3248·8	0·0
	350·0	2886·3	0·0	2776·6	0·0
	300·0	2391·9	0·0	2314·4	0·0
	250·0	1915·7	0·0	1865·7	0·0
	200·0	1462·7	0·0	1434·7	0·0
	180·0	1288·1	+0·1	1267·7	0·0
	160·0	1118·7	0·0	1104·0	0·0
	140·0	954·7	0·0	944·8	0·0
	120·0	795·4	+0·1	790·9	0·0
	100·0	644·2	0·0	643·2	−0·2
	80·0	498·4	+0·1	498·9	0·0
	60·0	360·4	+0·1	362·3	+0·1
	40·0	231·1	0·0	233·3	0·0
	20·0	110·7	+0·1	112·5	−0·1
	0·0	0·0	0·0	0·0	0·0

* The temperature differences are (observation − reference table).

It was mentioned earlier in this paper that the platinum, 10% rhodium-platinum thermocouple is the standard interpolation instrument for realizing the International Temperature Scale over the range from 630·5° c. to 1,063° c. It is of interest to compare the results of interpolated values obtained by the specified method using the quadratic equation based on the E.M.F. values at the freezing points of antimony, silver and gold with those given by the interpolation methods described in the Appendix. Taking the values of 5,539·5, 9,137·9 and 10,325·3 μv. for the temperatures 630·5°, 960·8° and 1,063° c. as given in Table 1 for the 10% thermocouples, the following equation is derived:

$$E = -315{\cdot}3 + 8{\cdot}2305t + 0{\cdot}001674t^2,$$

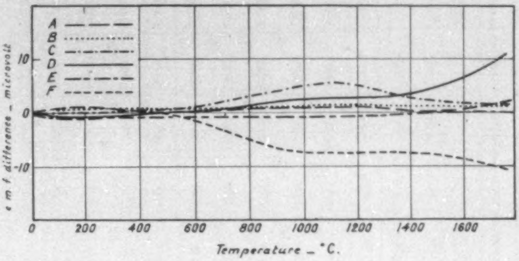


Figure 5. Departures of E.M.F. of each couple from mean of six 13% couples ot make A.

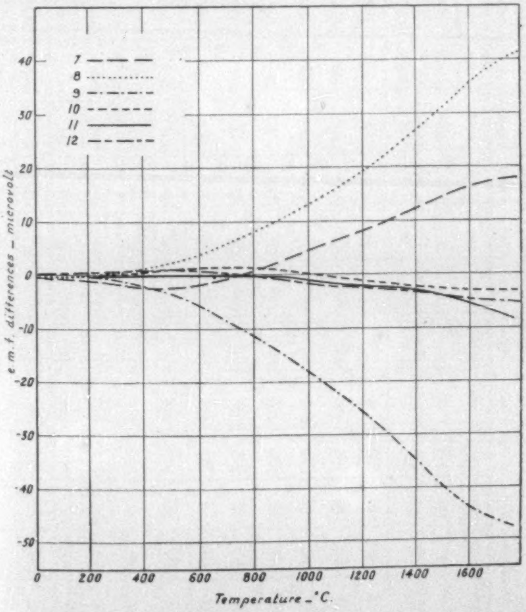


Figure 6. Departures of E.M.F. of each couple from mean of six 13% couples of make B.

Table 2

Reference Table for Platinum, 10% Rhodium–platinum Thermocouples, 1949

This Table is based on the mean calibration of twelve thermocouples, six from each of two manufacturers. The calibration was obtained by comparison with the resistance thermometer, by observations at the freezing points of silver, gold and platinum and by comparison with the optical pyrometer. The temperature scale used is the International Temperature Scale 1948 which, in the range covered by the Table, is based on the following fixed points :—

Melting point of ice	0° c.	Freezing point of silver	960·8° c.
Boiling point of water	100° c.	Freezing point of gold	1,063° c.
Boiling point of sulphur	444·60° c.		

and from 1,063° c. upwards on the Planck law of radiation with a value of 1·438 cm. degrees for the constant C_2.

Temperature in ° c. (Int. 1948). E.M.F. in microvolts (absolute). Cold junctions at 0° c.

t(°c.)	0	100	200	300	400	500	600	700	800	900	1000	1100	1200	1300	1400	1500	1600	1700	t(°c.)
0	0 55	642 74	1435 85	2314 92	3249 95	4220 99	5226 103	6265 106	7339 109	8446 113	9591 116	10757 118	11946 120	13155 122	14371 121	15580 120	16770 117	17922 113	0
10	55 57	716 75	1520 85	2406 92	3344 96	4319 100	5329 103	6371 106	7448 110	8559 113	9707 116	10875 118	12066 121	13277 122	14492 121	15700 120	16887 117	18035 112	10
20	112 60	791 76	1605 86	2498 92	3410 97	4419 100	5432 103	6477 107	7558 110	8672 114	9823 116	10993 118	12187 120	13399 121	14613 121	15820 120	17004 116	18147 112	20
30	172 61	867 78	1691 87	2590 93	3537 96	4519 100	5535 103	6584 107	7668 110	8786 114	9939 116	11111 118	12307 121	13520 122	14734 121	15940 120	17120 116	18259 112	30
40	233 64	945 79	1778 88	2683 94	3633 97	4619 100	5638 104	6691 107	7778 110	8900 114	10055 117	11229 119	12428 121	13642 122	14855 121	16060 119	17236 115	18371 111	40
50	297 66	1024 80	1866 88	2777 94	3730 97	4719 101	5742 104	6798 108	7888 111	9014 115	10172 117	11348 119	12549 121	13764 122	14976 121	16179 119	17351 115	18482 110	50
60	363 67	1104 81	1954 89	2871 94	3827 98	4820 101	5846 104	6906 108	7999 111	9129 115	10289 117	11467 119	12670 121	13886 121	15097 121	16298 118	17466 115	18592	60
70	430 69	1185 83	2043 90	2965 94	3925 98	4921 101	5950 105	7014 108	8110 112	9244 115	10406 117	11586 120	12791 121	14007 121	15218 121	16416 118	17581 114		70
80	499 71	1268 83	2133 90	3059 95	4023 98	5022 102	6055 105	7122 108	8222 112	9359 116	10523 117	11706 120	12912 122	14128 122	15339 120	16534 118	17695 114		80
90	570 72	1351 84	2223 91	3154 95	4121 99	5124 102	6160 105	7230 109	8334 112	9475 116	10640 117	11826 120	13034 121	14250 121	15459 121	16652 118	17809 113		90
100	642	1435	2314	3249	4220	5226	6265	7339	8446	9591	10757	11946	13155	14371	15580	16770	17922		100

Table 3

Reference Table for Platinum, 13% Rhodium–platinum Thermocouples, 1949

This Table is based on the mean calibration of twelve thermocouples, six from each of two manufacturers. The calibration was obtained by comparison with the resistance thermometer, by observations at the freezing points of silver, gold and platinum and by comparison with the optical pyrometer. The temperature scale used is the International Temperature Scale 1948 which, in the range covered by the Table, is based on the following fixed points :—

Melting point of ice	0° c.	Freezing point of silver	960·8 °c.
Boiling point of water	100° c.	Freezing point of gold	1,063° c.
Boiling point of sulphur	444·60° c.		

and from 1,063° c. upwards on the Planck law of radiation with a value of 1·438 cm. degrees for the constant C_2.

Temperature in ° c. (Int. 1948). E.M.F. in microvolts (absolute). Cold junctions at 0° c.

t(°c.)	0	100	200	300	400	500	600	700	800	900	1000	1100	1200	1300	1400	1500	1600	1700	t(°c.)
0	0 54	644 75	1463 89	2392 97	3397 104	4460 109	5571 114	6735 120	7952 124	9209 128	10510 133	11850 136	13222 139	14617 141	16039 143	17463 141	18855 137	20202 132	0
10	54 57	719 77	1552 89	2489 98	3501 104	4569 110	5685 114	6855 120	8076 124	9337 128	10643 133	11986 136	13361 139	14758 141	16182 143	17604 141	18992 136	20334 132	10
20	111 59	796 79	1641 91	2587 99	3605 105	4679 110	5799 115	6975 120	8200 125	9465 129	10776 133	12122 136	13500 139	14899 142	16325 143	17745 140	19128 136	20466 132	20
30	170 61	875 80	1732 91	2686 100	3710 105	4789 110	5914 116	7095 121	8325 125	9594 129	10909 133	12258 137	13639 139	15041 142	16468 142	17885 140	19264 135	20598 131	30
40	231 64	955 81	1823 93	2786 100	3815 106	4899 111	6030 116	7216 122	8450 126	9723 130	11042 134	12395 137	13778 139	15183 142	16610 143	18025 140	19399 135	20729 131	40
50	295 66	1036 83	1916 94	2886 101	3921 107	5010 111	6146 117	7338 122	8576 126	9853 130	11176 134	12532 137	13917 140	15325 143	16753 142	18165 139	19534 134	20860 130	50
60	361 68	1119 84	2010 94	2987 102	4028 107	5121 112	6263 117	7460 122	8702 126	9983 131	11310 135	12669 138	14057 139	15468 142	16895 143	18304 138	19668 134	20990	60
70	429 70	1203 86	2104 95	3089 102	4135 108	5233 112	6380 118	7582 123	8828 127	10114 132	11445 135	12807 138	14196 140	15610 143	17038 142	18442 138	19802 134		70
80	499 72	1289 86	2199 96	3191 103	4243 108	5345 113	6498 118	7705 123	8955 127	10246 132	11580 135	12945 139	14336 140	15753 143	17180 142	18580 138	19936 133		80
90	571 73	1375 88	2295 97	3294 103	4351 109	5458 113	6616 119	7828 124	9082 127	10378 132	11715 135	13084 138	14476 141	15896 143	17322 141	18718 137	20069 133		90
100	644	1463	2392	3397	4460	5571	6735	7952	9209	10510	11850	13222	14617	16039	17463	18855	20202		100

which gives values of E.M.F. of 6,799, 7,340 and 7,890 μv. at 750°, 800° and 850° c. respectively. These are in excess of the values in Table 1 by only 1, 1 and 2 μv. respectively.

Another point of interest is to observe the measure of agreement given by the tables with the International Temperature Scale criteria for the 10% type of couple to establish its suitability for use as a standard instrument. The criteria are

$$E_{\mathrm{Au}} - E_{\mathrm{Ag}} = 1{,}185 + 0{\cdot}158(E_{\mathrm{Au}} - 10{,}310) \pm 3\,\mu\mathrm{v}.$$

and

$$E_{\mathrm{Au}} - E_{\mathrm{Sb}} = 4{,}776 + 0{\cdot}631(E_{\mathrm{Au}} - 10{,}310) \pm 5\,\mu\mathrm{v}.,$$

where E_{Au}, E_{Ag} and E_{Sb} are the E.M.F.'s in microvolts at the freezing points of gold, silver and antimony. These criteria yield values on the basis of $E_{\mathrm{Au}} = 10{,}325{\cdot}3$ from Table 1 of 1,187·2 and 4,784·9, and the tables give 1,185·9 and 4,783·9, the agreement being well within the prescribed limits.

In the introduction to this paper it was mentioned that the reference table for 13% couples over the range from 1,400° to 1,750° c. was published in 1943, and at that time a full comparison of optical pyrometer and thermocouples could not be made. When this table has been corrected to the 1948 temperature scale and changed to absolute microvolts, the differences from the values in Table 3 shown in Figure 7 are obtained. The 1943 table was based on thermocouples having a mean E.M.F. at the palladium point of 18,198 absolute microvolts. The corresponding value for the present tables is 18,193, so that the tables refer to thermocouples of approximately the same E.M.F. The differences shown in Figure 7 nowhere exceed the equivalent of 1° c.

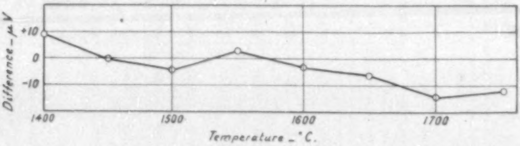


Figure 7. Difference between present reference table and 1943 table. (Ref.: *J. Iron Steel Inst.*, No. 1, 1943, p. 205.)

ACKNOWLEDGMENTS

The work described above has been carried out as part of the research programme of the National Physical Laboratory, and this paper is published by permission of the Director of the Laboratory. The author desires to acknowledge the assistance rendered by Mr. L. H. Pemberton who carried out much of the observational work and by Mr. J. A. Hall for comments and suggestions.

REFERENCES

Barber, C. R., 1943, *J. Iron Steel Inst.*, No. 1, 205.

Comité International des Poids et Mesures, 1948, *Échelle Internationale de Température, Procès-Verbaux des Séances*, t. xxi.

Grace, A., and Hall, J. A., 1943, *J. Sci. Instrum.*, **20**, 60.

APPENDIX

Computation of the Reference Tables by L. Fox

Mathematics Division, National Physical Laboratory

There is no simple standard method for producing a smooth table, linearly interpolable, from a set of quantities obtained by physical measurement and given at unequal intervals of the argument. The lack of statistical knowledge of the conditions of the experiments, the weights to be attached to the various observations, and a probable mathematical relation between the functions concerned suggest that graduation in the actuarial and statistical sense is here impossible and even superfluous. The only criterion is that the final table should be smooth, and that the tabular entries corresponding to the arguments of the given data should differ as little as possible from the observed quantities.

The method adopted here is based on this criterion, and consisted of four stages:

(i) From standard processes of interpolation, using differences and divided differences where necessary, values of the E.M.F. were obtained at equal intervals of 50° C. over the whole temperature range. The differences will not be smooth, but after a certain order will oscillate about a mean value which is approximately zero. Differences of orders higher than this are neglected in the interpolation process. In the present case the first neglected difference was the fourth.

(ii) The values so obtained at equal intervals were differenced, and again the differences of some order showed oscillation about zero. This difference was smoothed, by a kind of relaxation process, making suitable changes in the functional values to reduce the differences to very small numbers oscillating about zero. Some skill is required here to produce the desired effect with the minimum functional change.

(iii) The smoothed values were then interpolated to fifths on the National Accounting Machine, the resulting tables having an argument interval of 10° C. and being linearly interpolable.

(iv) As a final check, the temperature arguments corresponding to the observed voltages were obtained by linear inverse interpolation and compared with the experimental figures. The discrepancies are shown in Table 1, and are seen to lie well within the possible limits of experimental error. This provides a complete verification of the adequacy of the process.

The work described above was carried out in the Mathematics Division of the National Physical Laboratory.

A Joule-Thomson Cascade Liquefier for Helium

By J. ASHMEAD
Royal Society Mond Laboratory, Cambridge

Communicated by the Low-Temperature Group, read at Group Meeting on 7th December 1949;
MS. received 23rd January 1950.

ABSTRACT. A helium liquefier has been constructed with an output of 3·8 litres per hour and a starting time of 70 minutes. The liquid hydrogen needed to cool the helium below its inversion temperature is provided continuously by a hydrogen liquefier within the same machine. The advantages of the machine are (i) short starting time, (ii) simplicity of operation and ease of servicing, (iii) closed hydrogen circuit.

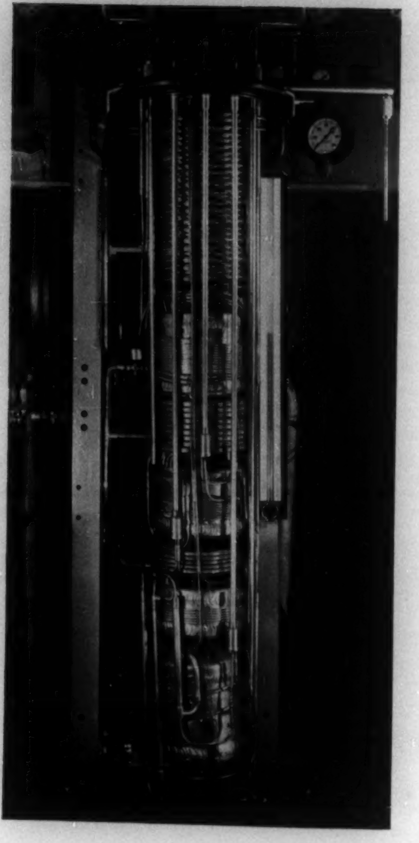
§ 1. INTRODUCTION

BEFORE the subject of the present paper is discussed, it seems desirable to put on record the performance of its predecessor in the Mond Laboratory, the liquefier built by Kapitza (1934). Exact data for the years before the war are not readily available, but it is known that up to September 1939, the liquefier had been used on more than 350 days, and probably produced about 2,000 litres of liquid helium. Between December 1945 and May 1949, 216 runs were made and helium for 554 experiments was produced. This is a very fine record for a machine which was really intended as an experimental prototype rather than a final model. The starting time of the liquefier (using, of course, pre-cooling with liquid nitrogen) was only 50 minutes, and the rate of making liquid was 1·7 litres per hour.

By 1947, it became clear that the larger cryostats and more ambitious research programmes of the post-war years made a liquefier of about twice the existing capacity desirable. Since careful consideration was given to the possible alternatives, it may be useful to record the reasons which led to the solution adopted, since some other laboratories may be influenced by similar considerations. The new liquefier had to satisfy as many as possible of the following conditions:

(i) Its output should be about 4 litres per hour.
(ii) If possible, the existing compressors and gas-holders should be used.
(iii) The apparatus should be as safe as possible.
(iv) The starting time should be short.
(v) Simplicity in running and ease in servicing are desirable even at the expense of a slight reduction in efficiency or an increase in the consumption of liquid nitrogen.

At first, the natural tendency was towards a re-designed liquefier of the expansion engine type. However, it appeared quite clear that no great increase in output was to be expected as long as the 'through-put' of the existing compressor was used. All that could be done was to produce a machine which might be a little more efficient than the Kapitza machine, and would be a good deal easier to service. To obtain the desired output, either a larger compressor would have to be installed, or else extra cooling would have to be provided by a subsidiary hydrogen circuit, as in the Moscow Kapitza machine.



In common with many other cryogenic laboratories the Mond Laboratory possesses a hydrogen liquefier, with a compressor and arrangements for storing the pure gas. It was therefore reasonable to consider at this stage those methods of helium liquefaction which demanded a supply of liquid hydrogen. The high thermodynamic efficiency of a liquefier of the Simon expansion type makes it very attractive in those cases where all the helium required for a day's experimenting can be produced in a single expansion, but if a number of successive compressions and expansions have to be made to provide all the liquid wanted, it seems less certain whether the *average* rate of production of liquid will be higher than with a continuously running, though less efficient, machine. Moreover, the Mond Laboratory does not possess that experience of small Simon liquefiers which other laboratories enjoy, so that the design of a large machine of that type would have been rather more of a new departure than most of the other possible devices.

The other important existing type of helium liquefier is that employing straightforward Joule–Thomson expansion of helium pre-cooled by liquid hydrogen, such as that in the Kamerlingh Onnes Laboratory at Leiden. Examination of the performance of this machine showed that it should be easy to obtain the output required with the helium compressor available. On the other hand, such a scheme had two important disadvantages. It involved the use of liquid hydrogen, with its attendant dangers, and it entailed two starting times instead of one. The second of these disadvantages could be overcome by making the hydrogen on a previous day and storing it (see for example Jones, Larsen and Simon 1948), so that the liquid helium could be produced fairly easily on the days when it was needed, but it seemed as though there might be another and more attractive possibility.

The suggestion was that a hydrogen and helium liquefier should be combined within the same container. The liquid hydrogen produced would never be decanted from the machine, but would serve only to cool the helium below its inversion temperature. A small liquefier of this type was described by Seiler (1941) but it differs in so many respects from the present machine that the two are hardly comparable. The advantages of a liquefier of this type are appreciable. The hydrogen circuit is completely closed and the gas therefore needs purification only very infrequently. Further, since no liquid hydrogen ever leaves the apparatus, the explosion hazard is considerably reduced compared with a system in which liquid or gas is frequently handled. Again, the combination of the two liquefiers in one envelope enables a very good overall starting time to be obtained —certainly less than the sum of the two separate starting times. Experience with the existing hydrogen liquefier led us to believe that a liquefier of the Joule–Thomson type would be straightforward to run. These arguments resulted in the choice of the type of liquefier described in this paper, and subsequent experience of the completed machine has, in our opinion, justified the decision made.

§ 2. GENERAL DESIGN OF THE LIQUEFIER

Before the details of the various components of the machine are examined, it is convenient to describe the general arrangement of the liquefier illustrated in the Plate. The distance from the plate near the top of the liquefier to the floor is about five feet. High pressure hydrogen (150 atmospheres) and high pressure helium (25 atmospheres) are cooled in the top exchanger by returning hydrogen, returning helium, and nitrogen from the liquid nitrogen bottle just below the exchanger. The liquid nitrogen in this bottle boils at about 67° K. under reduced pressure and the high pressure hydrogen and helium pass through tubes soldered to the outside of the bottle and leave it at a temperature within one or two degrees of its own. In the next exchanger the high pressure gases are cooled by the returning gases, and after this exchanger the hydrogen expands at a Joule–Thomson valve to a pressure of 1·5 atmospheres. Part of the hydrogen is liquefied and collects in the upper hydrogen bottle and the remainder of the gas passes back up the heat interchangers. A small fraction (about one-fifth) of the liquid passes through a transfer valve into the lower hydrogen bottle where it boils at 6 cm. pressure, producing a temperature of 14° K. The gas from this lower bottle passes only through the short exchanger situated between the two hydrogen bottles, and then goes straight to room temperature via a wide-bore tube. This causes a slight loss of thermal efficiency but solves the problem of producing a high pumping speed at the lower bottle. Two hydrogen bottles are used rather than one because if all the gas were expanded to 6 cm. pressure, the speed of the vacuum pump would have to be about 25 times as great. The high pressure helium is cooled by passage through the tubes soldered to the outside of both hydrogen bottles, and by the hydrogen gas in the intermediate exchanger.

The helium thus enters the final exchanger at 14° K., and after passing through the exchanger it expands at a Joule–Thomson valve where part of it liquefies and collects in the large copper bottle at the bottom of the machine. From there, the liquid is siphoned into the flasks through the vacuum jacketed tube which can be seen near the top right-hand corner of the photograph.

An important feature of the design of this machine lies in the precautions taken to reduce the starting time. During the early stages of operation the hydrogen and helium do not return up the exchangers but pass straight from the bottles back to the gas-holders via 'circulation valves' on the top of the liquefier. This causes a certain waste of cold in the top exchanger, and hence a greater consumption of liquid nitrogen, but it results in a more rapid cooling of those parts of the machine below the liquid nitrogen bottle. The wide-bore tubes to the circulation valves from the hydrogen and helium bottles are also connected via safety valves to the gas-holders in case a dangerous pressure builds up in either bottle.

Some of the more detailed features of the design may now be examined.

§ 3. THE HEAT INTERCHANGERS

The top three heat exchangers call for little comment. They consist essentially of bundles of tubes soldered together. The data for their design are taken from Jacobs and Collins (1940), and the actual method of construction is very similar to that employed by Blanchard and Bittner (1942). The bottom exchanger, on the other hand, is rather novel in design. At very low temperatures the high density of the helium allows it to flow turbulently even through very small bore tubes. An ideal heat exchanger would therefore consist, on the low pressure side, of a bundle of some fifty tubes of 1mm. bore, all in parallel. This effect can be produced by soldering to the high pressure tube a piece of crinkled copper foil, which is then wrapped round on itself and pushed inside a thin-walled

German silver tube. Thus the spaces between the crinkles form fine bore tubes giving very efficient heat transfer between the gas and the copper. The copper is near the maximum of its thermal conductivity in this temperature range, and so the cold is efficiently conducted to the central tube. To prevent longitudinal conduction, the copper foil is cut into pieces 6 cm. long. The form eventually taken by this heat interchanger is rather like a multiple trombone, and the first few sections of it can be seen surrounding the helium bottle at the left of the photograph.

§ 4. THE VALVES

The Joule–Thomson valves and the transfer valve are almost identical in design. The threaded part of the valve is at room temperature, near the top plate rather than close to the expansion orifice, so as to reduce the possibility of a valve seizing up because of impurities. The valves are so constructed that the whole valve, including the seating, can be withdrawn through the top plate after only two simple soldered joints have been broken. The design adopted for the circulation valves is a development from the tap used in the Kapitza liquefier for drawing off the liquid helium (Kapitza 1934).

§ 5. THE SIPHON TUBE

The liquid helium is drawn off from the helium bottle through a German silver tube over two metres long, 3 mm. in diameter and 0·1 mm. in wall thickness. This tube is, of course, protected by a vacuum jacket and a copper radiation shield, which is thermally bonded to the helium return tubes near the liquid nitrogen bottle.

§ 6. RADIATION SHIELD

The part of the liquefier below the nitrogen bottle is protected from the influx of room temperature radiation by a copper shield about 4 mm. thick, which is cooled by conduction from a subsidiary nitrogen bottle at its lower end.

§ 7. CASING

The liquefier is surrounded by a brass case with a Gaco ring seal to the brass plate from which the liquefier hangs. A vacuum of better than 10^{-5} mm. of mercury is maintained inside the casing by a diffusion pump of speed 20 litres per second. Even when the casing is shut off from the pump by a high vacuum valve, the thermal isolation is very good. After the liquefier is closed down, although no more liquid nitrogen is siphoned into the shield bottle, the rate of evaporation of helium is only 0·1 litre per hour and there is always some liquid hydrogen left in the hydrogen bottle the next day.

§ 8. INDICATORS

Besides the usual complement of pressure gauges, the liquefier is equipped with simple gas thermometers at the hydrogen and helium expansion valves. The levels of the liquids in the various bottles are observed by measuring on an inclined butyl phthalate manometer the pressure difference between the gas above the free surface in the bottle, and that in a tube connected to the bottom of the bottle. This second tube is curled in a horizontal plane just below the bottle, so that the free surface in the tube is always in the horizontal part of the tube, even if the heat flow down the tube varies appreciably. The butyl phthalate gauges are fitted with special self-returning safety traps, to guard against sudden variations of pressure, and they are by-passed by suitable valves until conditions are steady. Despite the very small pressure differences available owing to the low density of liquid hydrogen, the level indicators work very satisfactorily.

§ 9. GENERAL

A great deal of attention has been paid in the design of this machine to ease of servicing, and no attempt has been made to make the design compact by enclosing one part of the liquefier inside another. This principle has produced a rather tall machine, but one in which every vital part can be easily exposed for examination and repair. A pit has been dug in the floor under the liquefier so that the outer casing and the radiation shield can be removed with the minimum of trouble. About 20 minutes is required to remove the case and half an hour to disconnect the radiation shield, and very little longer to replace them. All the components hang from the top plate of the machine, which is supported by a cantilever so that everything is accessible.

§ 10. PERFORMANCE

Liquid hydrogen is produced about 30 minutes after the machine has started, and in a further 40 minutes the liquefier has reached equilibrium and is producing liquid helium. During this starting period 15 kg. of nitrogen is used in the main bottle and 10 kg. to cool down the shield. Once the shield is cold, very little more nitrogen is needed to keep it cold. In the equilibrium state, the through-put of hydrogen is 21 $m^3/hour$ at a pressure of 150 atmospheres, and the helium through-put is 17 $m^3/hour$ at 25 atmospheres. The consumption of liquid nitrogen is 8 kg/hour and the output of liquid helium is 3·8 litre/hour. If the machine is operated as a hydrogen liquefier, it produces 5 litre/hour, which is about 80% of the theoretical yield. The helium part of the circuit usually gives about 75% of the theoretical optimum, though as much as 85% has sometimes been obtained. The lower efficiency is accompanied by an increased pressure drop, and is probably due to a slight 'plating' of the exchangers by residual impurities.

ACKNOWLEDGMENTS

During the earlier stages of the design, I had many helpful discussions with Professor J. F. Allen, and Dr. D. Shoenberg encouraged me during the later phases of the work. A particular acknowledgment is due to Mr. F. Sadler who freely placed his experience of earlier machines at my disposal and was responsible for much of the detailed design work. The greater part of the actual construction was carried out most efficiently by Mr. E. Boud.

REFERENCES

Blanchard, E. R., and Bittner, H. W., 1942, *Rev. Sci. Instrum.*, **13**, 394.
Jacobs, R. B., and Collins, S. C., 1940, *J. Appl. Phys.*, **11**, 491.
Jones, G. O., Larsen, A. H., and Simon, F. E., 1948, *Research*, **1**, 420.
Kapitza, P., 1934, *Proc. Roy. Soc.* A, **147**, 189.
Seiler, K., 1941, *Ann. Phys., Lpz.*, **39**, 129.

Time Decrease of Magnetic Permeability in Alnico

By R. STREET and J. C. WOOLLEY
The University, Nottingham

Communicated by L. F. Bates; MS. received 23rd November 1949

ABSTRACT. The formal theory, based on the concept of thermal activation of domain processes, previously proposed to account for the phenomenon of magnetic viscosity, has been extended to describe time decrease of permeability. Within certain limits the predictions of the theoretical analysis are verified by measurements made on time decrease of permeability using an alnico rod specimen. It is shown that the quantitative results of these measurements are in good agreement with those obtained from the magnetic viscosity experiments. The possible physical mechanisms responsible for both magnetic viscosity and time decrease of permeability in alnico are discussed.

§ 1. INTRODUCTION

MEASUREMENTS of incremental permeability μ_i of a ferromagnetic material at any point on the magnetization curve are made by finding the change in magnetic induction ΔB produced by a known change in the magnetic field ΔH. For the majority of materials the value of μ_i so obtained is independent of the time which elapses (*resting time*) between the establishment of H and the application of the increment ΔH. However, when the incremental permeability of certain magnetic materials of low coercivity is examined in this way, μ_i is found to decrease as the resting time increases; this phenomenon is known as time decrease of permeability (Webb and Ford 1934). Snoek (1941) has shown that the effect may be observed in soft iron specimens only if they contain small quantities of carbon ($\sim$0·008%) or nitrogen ($\sim$0·007%) as impurity. He considers that the observed time decrease of permeability is due to a diffusion process of the impurity atoms within the iron lattice. Snoek has also investigated the magnetic viscosity exhibited by such soft iron specimens and has shown that there is an intimate relationship between the phenomena of magnetic viscosity and time decrease of permeability.

In this communication a report is given of the observations which have been made on time decrease of permeability in alnico, a high coercivity permanent magnet alloy. The related phenomenon of magnetic viscosity in alnico has already been investigated (Street and Woolley 1949 b, referred to below as I), and a formal theory of the phenomenon given. This theory was based on the concept of the activation of domain processes by thermal energy. In the first part of the present paper, this formal theory is developed to cover the case of time decrease of permeability, and discussion of the underlying physical mechanism is deferred until a later section.

§ 2. THEORY

As in I the underlying assumption is made that at some stage in the magnetization of a ferromagnetic substance a certain number of domains have their magnetization vectors in positions of metastable equilibrium. Hence, the domain vector orientation is scarcely altered by a small increase in the domain energy, but a greater energy increase initiates a transition to stable equilibrium which, in turn, involves a discontinuous, irreversible change in vector orientation. Consequently, the process envisaged is essentially one of activation, and the critical value of the energy required to produce discontinuous rotation will be referred to below as the activation energy of the domain. This energy can be supplied by an increase in the applied magnetic field in which case an increase in the magnetization of the specimen is observed. The experimental results given in I, which show that there is a spontaneous increase in the magnetization of the specimen in a steady applied field, are consistent with the assumption that the required activation energy may be supplied by random thermal fluctuations within the material. This leads to the prediction of time decrease of permeability. For, suppose that the applied field H (defined in § 1) is of such magnitude that spontaneous increase of magnetization occurs, then during the resting time t a certain number of domains are activated by thermal energy. Therefore, the number of domains which are activated by the eventual application of ΔH decreases as a function of t and, consequently, the corresponding value of ΔB decreases in a similar way, leading to an observed time decrease of permeability.

The quantitative analysis of this process is as follows. A steady field H is applied to the specimen at zero time. Let N be the number of domains having activation energies lying between E and $E+dE$ at time t.

Then,

$$\frac{dN}{dt} = -CN\exp\left(-\frac{E}{\boldsymbol{k}T}\right) \qquad \ldots\ldots(1)$$

where C is a constant, $\boldsymbol{k}$ is Boltzmann's constant and T is the absolute temperature of the specimen. In the derivation of this equation it is assumed that (*a*) successful activations do not produce additional metastable states with activation energies in the range from E to $E+dE$ and (*b*) each domain may be characterized by a constant activation energy independent of the state of the neighbouring domains, i.e. the magnetic interaction between domains is assumed to be negligible. Integrating equation (1) we have

$$\ln N = -\lambda t + \text{const.} \qquad \ldots\ldots(2)$$

where $\lambda = C\exp(-E/\boldsymbol{k}T)$.

In a general analysis, N should be taken as any arbitrary function of the activation energy, E. However, some definite form of this function is required for the further development of the analysis. In I the function was given the simple form defined as follows: (*a*) the domains have activation energies lying only in the range from zero to an upper limit, E_0, and (*b*) within this range, N at time $t=0$ is independent of E and has a value $p\,dE$, where p is the distribution constant. As far as the phenomenon of magnetic viscosity is concerned, the result of the theoretical analysis using this simplification is in very good agreement with our experimental observations. This form of the distribution is therefore retained in the following analysis of time decrease of permeability. Thus equation (2) becomes

$$N = p\exp(-\lambda t)\,dE. \qquad \ldots\ldots(3)$$

When ΔH is applied, the magnetization vectors of all domains having activation energies below some value E' will rotate discontinuously, the actual value of E' being directly determined by ΔH. The resultant total change in the intensity of magnetization of the specimen will consist of two parts, (*a*) the contribution of the discontinuous irreversible movements, which will be denoted by $\Delta I'$, and (*b*) the contribution of the reversible movements of the magnetization vectors of all the domains of the material. In order to determine the contribution $\Delta I'$ it is necessary to find the total number of domains, N_{tot}, having activation energies lying between zero and E' at time t after the field H has been established. This is obtained from equation (3):

$$N_{\mathrm{tot}} = \int_0^{E'} p \exp(-\lambda t)\, dE.$$

Thus, if $\bar{i}$ is the average value of the change in the resolved part of the domain magnetization along the direction of the applied field due to the activations, then

$$\Delta I' = \bar{i}p \int_0^{E'} \exp(-\lambda t)\, dE. \qquad \ldots\ldots(4)$$

Changing the variable to λt, equation (4) becomes

$$\begin{aligned}\Delta I' &= -\bar{i}p\mathbf{k}T \int_{Ct}^{\lambda' t} \frac{1}{\lambda t} \exp(-\lambda t)\, d(\lambda t)\\ &= \bar{i}p\mathbf{k}T[\mathrm{Ei}(-Ct) - \mathrm{Ei}(-\lambda' t)],\end{aligned} \qquad \ldots\ldots(5)$$

where $\lambda' = C \exp(-E'/\mathbf{k}T)$, and $-\mathrm{Ei}(-x)$ is the exponential integral, defined (Glaisher 1870) by

$$\int_x^{\infty} \frac{1}{u} \exp(-u)\, du.$$

The total change in the intensity of magnetization due to the application of ΔH will contain a contribution from reversible rotations of the domains, and this may be accounted for by the addition of a term $f(\Delta H)$ to the right-hand side of (5) which now becomes

$$\Delta I_{\mathrm{tot}} = \bar{i}p\mathbf{k}T[\mathrm{Ei}(-Ct) - \mathrm{Ei}(-\lambda' t)] + f(\Delta H). \qquad \ldots\ldots(6)$$

In the derivation of equation (6) it has not been necessary to make any assumptions concerning the ranges of t and E'. A general equation of very similar form to that of equation (6) has been derived in the case of magnetic viscosity (Street and Woolley 1949 c), viz.

$$\Delta I = \bar{i}p\mathbf{k}T[\mathrm{Ei}(-\lambda_0 t) - \mathrm{Ei}(-Ct) + E_0/\mathbf{k}T], \qquad \ldots\ldots(7)$$

where ΔI is the spontaneous increase in the intensity of magnetization after the resting time t, and $\lambda_0 = C \exp(-E_0/\mathbf{k}T)$. It has been found possible to make a direct check of equation (7) in its most general form since it is in satisfactory agreement with the experimental observations made by Snoek (1947) on magnetic viscosity in Mn–Zn ferrite.

A general plot of equation (7) shows that over a range of resting times t defined by the inequalities $t_0 > t > t_c$, where

$$\left.\begin{aligned}&|\mathrm{Ei}(-Ct_c)| \ll |\mathrm{Ei}(-\lambda_0 t_c)|\\ &\text{and} \qquad \lambda_0 t_0 \ll 1\end{aligned}\right\} \qquad \ldots\ldots(8)$$

the value of ΔI is given by

$$\Delta I = \bar{i}p\mathbf{k}T \ln t + \text{const.}$$

The convenient range of times during which experimental observations on magnetic viscosity in alnico were made was found to lie well within the limits of t defined immediately above. The experimental results were thus satisfactorily represented by the preceding $\ln t$ relation.

Subject to the limitations already imposed in the derivation of equation (6), the latter should give a satisfactory description of time decrease of permeability. Since the experimental results on this phenomenon given in the present paper were obtained for the same alnico specimen over the same time ranges which were employed in the magnetic viscosity experiments, (6) is now considered over a similarly limited range of times, viz. $t_0' > t > t_c'$ where

$$\left.\begin{array}{l} |\mathrm{Ei}(-Ct_c')| \ll |\mathrm{Ei}(-\lambda' t_c')| \\ \text{and} \qquad \lambda' t_0' \ll 1. \end{array}\right\} \qquad \ldots\ldots(9)$$

Within the range of times so specified equation (6) becomes

$$\Delta I_{\mathrm{tot}} = -\bar{i}p\mathbf{k}T(\ln \lambda' t + \text{const.}) + f(\Delta H). \qquad \ldots\ldots(10)$$

The limiting values of t_0' and t_c' defined by (9) now depend on the value of E', and hence on the corresponding value of the applied magnetic field increment ΔH. From the experiments on magnetic viscosity in alnico it would be expected that the time range from t_c' to t_0' would include the times conveniently available for experiment (from 30 to 3,000 sec.) when E' is comparable with E_0, i.e. when relatively large values of ΔH are used in the experiments.

In experimental work designed to test the predictions of equation (10) certain variables may be conveniently held constant. For example, if the increment in applied magnetic field, ΔH, is maintained at a constant value in a series of experiments, since E' is now constant equation (10) may be written

$$\Delta I_{\mathrm{tot}} = -\bar{i}p\mathbf{k}T(\ln t + \text{const.}) + f(\Delta H). \qquad \ldots\ldots(11)$$

Again, when the time t and the absolute temperature T are held constant, the variation of ΔI_{tot} as a function of E' is

$$\Delta I_{\mathrm{tot}} = \bar{i}pE' + \text{const.} + f(\Delta H). \qquad \ldots\ldots(12)$$

The present experimental work was directed towards the verification of equations (6), (11) and (12) and the evaluation of the constants characteristic of the domain processes involved in the magnetization cycle.

§ 3. APPARATUS AND METHOD

The rod of alnico employed in the present investigations was that used in I where details of its composition, dimensions and heat treatment are given. The specimen, on which a search coil was wound, was supported symmetrically either in an electrically heated tubular oven, the heater of which was wound non-inductively and fed with D.C. or, when temperatures below that of the room were required, in a Dewar vessel. The required external magnetic field was applied by means of a surrounding water-cooled solenoid capable of producing a maximum field of 1,100 oersteds. The current flowing in the solenoid, supplied by a 280 volt battery of accumulators, was measured by a potentiometer method. The search coil was connected to a ballistic galvanometer with a standard mutual inductance included in circuit for calibration purposes. Two galvanometers were used, one of period 6 seconds and the other of period 16 seconds.

The temperature of the specimen was determined by a calibrated thermocouple attached to its mid-point, and measurements were made at fixed temperatures within the range $-80°$ c. to $130°$ c., the maximum temperature fluctuations about the steady values being approximately $2°$ c.

The general principle underlying the measurements has already been outlined in § 1. In the majority of the experiments, the steady applied field H was equal to the coercive field ${}_IH_C$ of the specimen. Initially the alnico rod was brought into a cyclic state of magnetization and the apparatus was then allowed to reach thermal equilibrium with external demagnetizing field ${}_IH_C$ applied to the specimen. Immediately before a measurement was made one more complete hysteresis cycle was traversed. After the waiting period t seconds had elapsed, the field increment ΔH was applied, and the resulting change in magnetic induction was measured by means of the ballistic galvanometer. It was found most convenient to apply the field increment ΔH to the specimen by passing a current through an auxiliary solenoid wound over the main solenoid. In this way it was possible to obtain values of ΔH which were consistent and reproducible to within $\frac{1}{4}\%$, a much better reproducibility than was obtained by a sudden increase of the current in the main magnetizing solenoid.

§ 4. RESULTS AND INTERPRETATION

Measurements were first made in an attempt to verify equation (6). A typical set of results of variation of ΔI_m with t is plotted logarithmically in Figure 1.

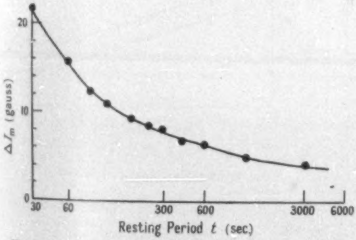


Figure 1. The variation of ΔI_m as a function of the resting period t. For each point the increment of the external field, $\Delta H = 6{\cdot}7$ oersteds. Galvanometer period = 6 sec.

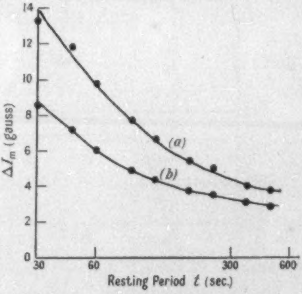


Figure 2. The variation of ΔI_m as a function of resting time, t sec., with $\Delta H = 4{\cdot}3$ oersteds.

Curve (*a*) galvanometer of period 16 sec.
Curve (*b*) galvanometer of period 6 sec.

ΔI_m is used to denote the measured value of the change of magnetization of the specimen. The measurements were made at room temperature and the increment ΔH was 6·7 oersteds for every individual determination. It was found that the curve drawn through these experimental points could not be represented by a difference of exponential integral functions as predicted by equation (6). This apparent failure of equation (6) to describe the experimental results is explained by the fact that, in its derivation, only those contributions to the measured change of magnetization directly due to the application of ΔH were considered. In practice, however, magnetic viscosity will occur at a field strength of $H + \Delta H$, i.e. immediately after the increment ΔH has been applied, and this will add a further contribution to the measured value of the change of induction because of the finite period of the ballistic galvanometer. Conclusive evidence of the existence of a magnetic viscosity contribution was obtained experimentally by making measurements on time decrease of permeability using two galvanometers with appreciably different periodic times, viz. 6 and 16 seconds. The results of measurements made at room temperature with $\Delta H = 4{\cdot}3$ oersteds are shown in Figure 2. These two sets of measurements were made under identical conditions except for the two different galvanometers used, and it is seen that the values of the changes in the intensity of magnetization obtained using the 16-second galvanometer (curve (*a*)) are greater than the corresponding values obtained using the 6-second period instrument. Thus the longer the periodic time of the galvanometer the greater the contribution of the magnetic viscosity to the measured change in intensity of magnetization.

In order to take account of this magnetic viscosity correction it is necessary to introduce another term into the right-hand side of equation (6). A consideration of equation (3), which gives the number of domains having activation energies lying between E and $E+dE$ as a function of E, shows the difficulty of estimating this correction term. The form of the required distribution functions may be calculated from equation (3), and the results of these calculations are shown diagrammatically in Figure 3, where N/p is plotted as a function of $E/\boldsymbol{k}T$ for the three values of Ct given for the individual curves. The domains represented by the shaded areas contribute to $\Delta I'$, whilst those represented by the unshaded areas are responsible for the disturbing magnetic viscosity effect. It is possible to obtain the analytical form of the distribution represented by the unshaded areas, but the resulting mathematical analysis of magnetic viscosity effects, assuming this form of distribution, is intractable, except for the case (*a*) of Figure 3, where the resultant distribution function at field $H+\Delta H$ has the simple form $N(E) = p = \text{constant}$. This corresponds to the case treated in I.

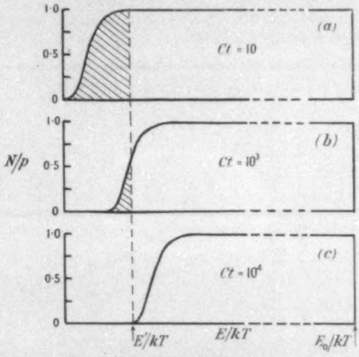


Figure 3. Variation of distribution function after various resting periods t. For explanation of shading and symbols see text.

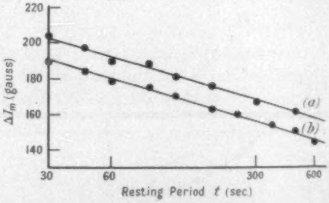


Figure 4. The variation of ΔI_m as a function of the resting time, t sec., with $\Delta H = 21{\cdot}4$ oersteds.

Curve (*a*) using galvanometer of period 16 sec. Curve (*b*) using galvanometer of period 6 sec.

Because of the limitations described above, an attempt was made to eliminate the contribution due to magnetic viscosity from the measured value of the change in induction by the following experimental method. A cascade of electromechanical relays was employed to disconnect the ballistic galvanometer from the search coil at various short times, δt (of the order of tens of milliseconds), after the increment ΔH had been applied. The observed change of induction for constant ΔH and constant resting period t was found to be a continuous function of δt. There was thus no justification for taking any particular value of the change of induction as corresponding to ΔI_{tot}. It was supposed that this result was due to (*a*) the time constant of the solenoid and battery circuit, and (*b*) the effects of eddy-current lag in the specimen, resulting in a relatively slow building up of the increment of field ΔH. Consequently, any calculation of the magnetic viscosity correction term which assumes an instantaneous application of ΔH is of little practical value.

Since the observed values of the change of induction contain an undetermined contribution due to magnetic viscosity, quantitative agreement between the experimental results and equation (6) cannot be expected. This difficulty is avoided to a considerable extent when, in any set of measurements, the contribution of magnetic viscosity is constant for all the individual measurements. It is shown below that this condition is satisfied if, after the application of the increment ΔH at time t, the *remaining* domains are characterized by a constant activation energy distribution p. For, from equation (1), the time rate of change of the intensity of magnetization at $t=0$ due to magnetic viscosity is given by

$$\begin{aligned}\left(\frac{dI}{dt}\right)_{t=0} &= \bar{i}pC\int_0^{E_0}\exp(-E/\boldsymbol{k}T)\,dE\\ &= \bar{i}p\boldsymbol{k}TC\left[1-\exp\left(-\frac{E_0}{\boldsymbol{k}T}\right)\right]. \qquad \ldots\ldots(13)\end{aligned}$$

In I it was shown that for this alnico specimen $E_0 \gg \boldsymbol{k}T$, a statement implicit in the first inequality of equation (8) above. Therefore, to a good approximation, equation (13) may be written as

$$\left(\frac{dI}{dt}\right)_{t=0} = \bar{i}p\boldsymbol{k}TC.$$

Thus the contribution to the measured change of magnetization due to magnetic viscosity in a time Δt, determined by the time of swing of the galvanometer, is approximately $\bar{i}p\boldsymbol{k}TC\,\Delta t$, which is a constant contribution for constant p. Reference to Figure 3 shows that the distribution function characteristic of the unactivated domains remaining after the application of ΔH will only be constant if the final conditions are as suggested by the curve (*a*). This means that for the maximum resting period to be used in the experiments the value of E', and consequently ΔH, must be of a sufficiently large magnitude. Thus for these cases equation (6) may be written

$$\Delta I_{\mathrm{m}} = \bar{i}p\boldsymbol{k}T[\mathrm{Ei}(-Ct)-\mathrm{Ei}(-\lambda' t)] + \text{const.} + f(\Delta H), \qquad \ldots\ldots(14)$$

where ΔI_{m} denotes the value of the change of magnetization which should be obtained in experimental measurements. The effect of magnetic viscosity is now included in the constant term on the right-hand side of equation (14), and now this constant depends not only on the domain constants but also on the constants of the ballistic galvanometer and the associated circuits. The limitations of this analysis now become apparent, for, owing to the inclusion of the indeterminate constant term in equation (14), only the slopes of the experimental curves may be used to determine the fundamental constants of the domain processes.

If the magnitude of ΔH is sufficiently large so that equation (14) holds, then, although the measured values of the change in magnetization should be different if observations are made with two ballistic galvanometers of different periodic times, the difference should be constant and independent of the resting period t. Experimental evidence for this conclusion is presented in the two sets of curves shown in Figures 2 and 4. In the former case, where the size of the field increment ΔH is 4·3 oersteds, the difference between the measured values with ballistic galvanometers of periods 6 seconds and 16 seconds is a function of the resting time t. However, if the experiment is repeated for a larger value of ΔH, namely 21·4 oersteds, the two sets of results differ only by a constant value. It will also be seen from Figure 4 that, with $\Delta H = 21{\cdot}4$ oersteds, the inequalities (9) are satisfied, for a linear variation of $(\Delta I_m, \ln t)$ is obtained. The equation of this straight line is given by (11), after modification to take into account the contribution due to magnetic viscosity, i.e.

$$\Delta I_m = -\bar{i}p\mathbf{k}T(\ln t - \text{const.}) + f(\Delta H) + \text{const.} \qquad \ldots\ldots(15)$$

From equation (15) the slope of the $(\Delta I_m, \ln t)$ curve is $-\bar{i}p\mathbf{k}T$. The magnitude of the slope of the lines shown in Figure 4 is 14·7 gauss, measured at a temperature of 14° c., which differs by less than 1% from the value of $\bar{i}p\mathbf{k}T$ derived from the previous measurements on magnetic viscosity described in I.

The values of ΔI_m were determined as a function of time, using various fixed values of ΔH. These results are plotted in Figure 5 and the values of ΔH used are indicated on the individual curves. It is apparent from this figure that when $\Delta H \geqslant 18$ oersteds then (*a*) ΔI_m is proportional to $\ln t$ for values of t from 30 to 600 seconds, and (*b*) the slopes of the individual lines are equal and have a common value 14·7 gauss; (temperature 14° c.). This is in agreement with the prediction of equation (15) that the slope of the $(\Delta I_m, \ln t)$ curve, i.e. $\bar{i}p\mathbf{k}T$ should be independent of ΔH, for sufficiently large values of ΔH. Thus the

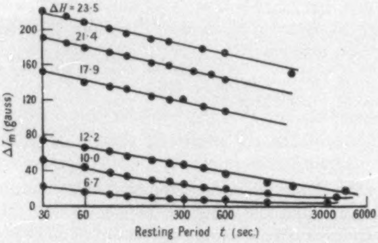


Figure 5. Variation of ΔI_m as a function of the resting period t for the various values of ΔH (oersteds) indicated on the curves. Galvanometer period = 6 sec.

experimentally determined slope is again in good agreement with the results of the magnetic viscosity measurements. The family of curves in Figure 4 serves to define the ranges of ΔH and t for which the inequalities given in equation (9) are justified.

A further prediction of equation (15) is that, provided $\overline{ip}$ is independent of temperature, the slope of the $(\Delta I_m, \ln t)$ curve should be proportional to the absolute temperature of the specimen. Experiments were made to check this prediction by making measurements of ΔI_m as a function of resting time t when the specimen was maintained at temperatures of −78°, 14°, and 130° c. The same value of ΔH, 21·4 oersteds, was used in the experiments at the three different temperatures, and the results are plotted in Figure 6. The slopes of the experimental curves have the values 9·9, 14·7 and 20·6 gauss. Thus the ratio of slope to absolute temperature, i.e. $\overline{ip}\boldsymbol{k}$, is constant, being equal to 0·051 gauss per degree Centigrade in each case, to within ±1%, giving a value of $\overline{ip}$ equal to $3{\cdot}7 \times 10^{14}$ gauss per erg, which is in good agreement with the value of $3{\cdot}6 \times 10^{14}$ gauss per erg derived from the measurements on magnetic viscosity given in I.

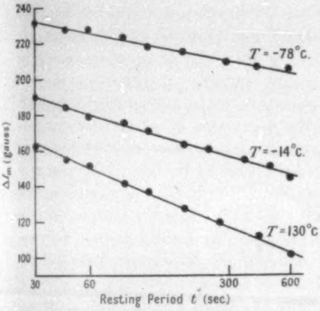


Figure 6. Variation of ΔI_m as a function of t with the specimen maintained at the different temperatures indicated on the curves, $\Delta H = 21{\cdot}4$ oersteds. Galvanometer period = 6 sec.

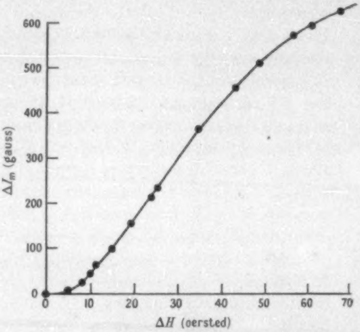


Figure 7. Variation of ΔI_m as a function of the change in the applied magnetic field ΔH for a constant resting period $t = 1$ minute. Galvanometer period = 6 sec.

The presence of the magnetic viscosity term precludes the possibility of determining the value of C from an analysis, based on equation (14), of the experimental results.

The experimental verification of equation (12) is not directly possible since the forms of $f(\Delta H)$, introduced to account for reversible changes, and $F(\Delta H) = E'$ are not known *ab initio*. However, measurements were made of ΔI_m as a function of ΔH with a resting period $t = 60$ seconds for each measurement; the temperature of the specimen was constant at 20° c. The results are plotted in Figure 7 and it will be seen that over a range of ΔH from 10 to 40 oersteds, ΔI_m is practically a linear function of ΔH. Qualitatively, with the constant distribution factor p assumed in the analysis, as indicated in Figure 3, it would be expected that $\Delta I'$, the contribution to the change in intensity of magnetization due to irreversible activation of the domains, would reach a constant value when $E'=E_0$. This corresponds physically to the situation when all the irreversible domain changes are activated by the application of ΔH. The consequence is that a decrease in the slope of the $(\Delta I_m, \Delta H)$ curve would be expected for values of ΔH above that corresponding to E_0. This is a possible explanation of the decrease of slope of the $(\Delta I_m, \Delta H)$ curve, for values of ΔH greater than 40 oersteds.

It is not possible to go further in the interpretation of the results shown in Figure 7 in terms of equation (12), without making further assumptions. The underlying physical mechanism responsible for the phenomena will now be considered.

§ 5. DISCUSSION OF POSSIBLE PHYSICAL MECHANISMS

The agreement obtained between the experimental results on time decrease of permeability and (*a*) the theoretical analysis above and (*b*) the experimental values obtained from magnetic viscosity measurements demonstrates the essential correctness of the activation concepts postulated to account for both phenomena. Any mechanism suggested must be compatible with the assumptions made in the development of the formal theory. One possible mechanism, already suggested in I, is based on the theory recently put forward by Stoner and Wohlfarth (1948) to account for the high values of the coercivity of permanent magnet alloys. These materials are supposed to consist of small ferromagnetic islands embodied in a less ferromagnetic matrix and the sizes of the islands are such as to preclude the formation of domain boundaries within them. It is shown that over certain ranges of the hysteresis cycles of such materials magnetization proceeds by discontinuous and irreversible rotations of the single domain vectors. In a simplified case, when the islands are prolate ellipsoids orientated at random, the discontinuous rotations occur over a range on the hysteresis cycle lying between a coercivity point and saturation. At any point in this range a certain number of domains have their magnetization vectors in positions of metastable equilibrium; that is, the free energy of a domain has a minimum value both before and after a discontinuous change of orientation, and transition from the initial to the final state may only occur if sufficient energy to overcome the intervening potential barrier is supplied. Thus Stoner's analysis would account for the potential barriers which may be overcome by thermal activation energy. In addition, as has already been shown in I, this theory would explain (*a*) the assumption that during the activation process no new domains of activation energy E are produced by successful activation processes, and (*b*) the experimentally observed fact that $\bar{ip}$ is independent of temperature.

Néel (1949 a) has examined the phenomenon of magnetic viscosity occurring in materials having a small quantity of ferromagnetic substance dispersed within them. He supposes that the ferromagnetic substance occurs in the form of particles of size below that necessary for domain boundary formation, i.e. single domains embedded in a non-ferromagnetic matrix are present. The postulate of thermal activation and the resulting $\log t$ relationship obtained by Néel are essentially similar to those advanced by the present authors and summarized in a note (Street and Woolley 1949 a). In addition, Néel has also suggested a mechanism whereby thermal deformation of the crystal lattice of the single domains may give rise to discontinuous rotations of the magnetization vectors

In a subsequent paper, Néel has considered the phenomenon of magnetic viscosity in 'massive' materials (1949 b), and has shown that displacements of the domain boundary walls, due to thermal activation, may occur. This is a possible mechanism to account for magnetic viscosity phenomena in massive materials, and Néel states that the general results of this analysis are similar to those obtained in the consideration of single domain materials. Further information on this work has not yet been published, but, as far as the results on alnico are concerned, it is difficult to see how such irreversible boundary displacements are consistent with a value of $\bar{i}p$ independent of temperature and the non-production of new metastable states after successful activations. Further, any proposed mechanism should be capable of giving a satisfactory account of the high coercivity values of alnico, and in this respect the theory of Stoner and Wohlfarth appears to be the more likely.

§6. EFFECT OF HEAT TREATMENT

Measurements of magnetic viscosity and time decrease of permeability have been made using alnico specimens which have been subjected to various heat treatments. The conclusion reached is that, if the heat treatment is such as to produce material of highest coercivity, then the magnetic viscosity effects are substantially reduced. It is hoped to publish further details of this work at a later date.

ACKNOWLEDGMENTS

The authors thank Professor L. F. Bates for the facilities of his laboratory and his continued interest and encouragement in this work, and Dr. K. Hoselitz for supplying the specimen used.

REFERENCES

Glaisher, J. W. L., 1870, *Phil. Trans. Roy. Soc.*, **160,** 367.

Néel, L., 1949 a, *Ann. Géophys.*, **5,** 99; 1949 b, *C. R. Acad. Sci., Paris,* **228,** 1210.

Snoek, J. L., 1938, *Physica,* **5,** 663; 1941, *Ibid.*, **8,** 711.

Stoner, E. C., and Wohlfarth, E. P., 1948, *Phil. Trans. Roy. Soc.*, **240,** 599.

Street, R., and Woolley, J. C., 1949 a, *Proc. Phys. Soc.* B, **62,** 141; 1949 b, *Proc. Phys. Soc.* A, **62,** 562; 1949 c, *Ibid.*, **62,** 743.

Webb, C. E., and Ford, L. A., 1934, *J. Instn. Elect. Engrs.*, **75,** 787.

The Magnetic Properties of Uranium and Uranium–Iron Alloys

By L. F. BATES and J. R. MALLARD
University of Nottingham

MS. received 6th January 1950

ABSTRACT. Measurements have been made of the magnetic susceptibility of fairly pure uranium and of 5 and 10 atomic per cent alloys of iron with uranium. Using a modification of the Curie method, with special pole-pieces and electromagnetic dynamometer control, the specimens were found to be paramagnetic with no trace of ferromagnetism in the temperature range 20° to 350° c. At 20° c. the susceptibility of pure uranium is $+1{\cdot}740\times10^{-6}$ E.M.U. per gm., increasing to $+1{\cdot}804\times10^{-6}$ at 350° c.; the results fit the equation

$$\chi_U=32\times10^{-11}T+1{\cdot}564\times10^{-6}+24{\cdot}0\times10^{-6}/T.$$

Assuming that the iron is present in form of the compound $U_6.Fe$, it is found that the latter has a susceptibility of $+2{\cdot}036\times10^{-6}$ E.M.U. per gm. at 20° c., which remains practically constant with temperature.

---

§ 1. INTRODUCTION

In the last of a series of papers on the magnetic properties of metals dissolved in mercury, referred to hereinafter as A, Bates and Somekh (1944) described an apparatus for measuring the magnetic susceptibility of non-homogeneous systems and its use in studying cadmium, gold, silver and zinc amalgams over wide ranges of concentrations at room temperature. In this work ferromagnetic impurities were rigorously avoided, so that little or no variations of field conditions were required. As the Curie method was used, the amalgams were enclosed in one of two long cylindrical tubes mounted so that they could move along magnetic lines of force running between special cylindrical pole pieces in such a way that the force exerted on each element of amalgam was as nearly as possible independent of its position in the tube (cf. Bates 1948). Details of the main features of the apparatus and the basic theory of measurement by a dynamometer method are given in A, pp. 183 to 185.

It was decided, in view of its present-day importance, to measure the susceptibility of uranium, kindly supplied by the Atomic Energy Research Establishment at Harwell, with the above apparatus improved in certain particulars as described below. It was clearly necessary to deal with possible effects of ferromagnetic impurities and also to make measurements above room temperature. The specimen tube was therefore mounted inside a vertical, cylindrical electric furnace consisting of a copper tube wrapped with asbestos paper, wound non-inductively with nichrome strip, and coated with ‘purimachos’ cement. The whole was placed inside a brass cooling jacket through which water was continuously circulated. A copper–constantan thermocouple with one junction placed directly below the specimen tube indicated the temperature; the leads to the junction passed through holes drilled in a ‘sindanyo’ base to which the jacket was fixed.

In A, the specimen tubes were of such a length that the centre of the specimen was located some 18 cm. below the glass hook of the stopper, the latter being hooked on to the copper framework extension of the electrodynamometer with a touch of paraffin wax melted upon the junction. When the furnace temperatures rose above 100 to 150° C. this wax joint melted and the tube sometimes changed its position. Accordingly, in both the glass hook and the copper hook a well defined V notch was filed, and a small length of plaited silk, with a loop at each end, was used to connect the hooks and suspend the tube. Lateral displacements due to temperature changes were thus effectively eliminated and the tube could hang vertically with the stopper in a standard position, whether empty or containing liquid or solid specimens which fitted the tube. Of course, solids which did not exactly fit the tube might displace it from the vertical and so alter the effective length of the couple arm of the electrodynamometer. However, by repeating the measurements in such a case with the tube carefully rotated through 180° so that the displacement is reversed, this defect could, clearly, be overcome. The silk connection also had the unexpected advantage that accidental collisions of the tube with the furnace walls did not cause trouble.

The apparatus was calibrated with mercury distilled by the Hulett method in an all-glass apparatus to reduce contamination; the susceptibility of mercury was taken as $-0{\cdot}1680 \times 10^{-6}$ E.M.U. per gm. The specimen tubes were evacuated by the simple method described in A. A new electrodynamometer was constructed, with the coil wound on a former in order to give less distortion and to deal with larger couples.

The A.C. current to the dynamometer coils was supplied by a constant voltage transformer delivering 230 volts r.m.s. at 500 watts, the voltage fluctuations being less than 0·2%. The same electromagnet was used as in A, but it was supplied from a battery of heavy-duty accumulators.

§ 2. EXPERIMENTAL MEASUREMENTS

For preliminary tests a specimen of pure uranium was turned down from a small rod, taking the recommended health precautions during the turning and handling, to form a cylinder 1·5 cm. long and 5 mm. diameter weighing about 5 gm. To prevent rapid oxidation of the freshly turned surface the uranium was dropped into a dish of concentrated nitric acid—which dissolves the oxide but only attacks the metal slowly—quickly transferred to a dish of absolute alcohol and washed thoroughly in the latter. After rapid drying with filter paper it was slipped into the specimen tube, which was then evacuated rapidly. Less speed in manipulation was necessary with iron–uranium alloys which oxidize much more slowly.

The pure uranium was tested for ferromagnetism by measuring the susceptibility k_H for different fields H and plotting k_H as a function of $1/H$ (cf. Bates 1948, p. 114); if no impurity is behaving as a ferromagnetic substance the graph should be parallel to the $1/H$ axis. The values of H were small enough to ensure that the pole-pieces were far from saturated, so that the magnetic field was always of the form $H(dH/dr) = \text{const.}$ Special tests were also made to ensure that for a given value of H equal portions of the specimen were all exposed to the same magnetic force. This was done by measuring the forces on different volumes of liquid in the specimen tube and showing how the force depended on the volume occupied. As a final result of these tests, the specimen tube was never occupied to a height greater than 3 cm.; the occupied portion was kept as closely as possible midway between the upper and lower surfaces of the pole-pieces, and a field current of 3·6 amp. was not exceeded. This reduction in specimen length was compensated by using new specimen tubes of larger internal diameter.

For final measurements on pure uranium at room temperature a rod 1·35 cm. long closely fitting the specimen tube was used, and the restoring couples were found for some six suitable values of the field current. For a given value of the field current, the two restoring couples, with the specimen tube in a chosen position and in another turned through 180° with respect to the first, differed by less than 0·5%. Couples on a specimen of pure mercury were measured in the same way; so, too, were those on the empty specimen tube.

Let χ_s and χ_{Hg} be the mass susceptibilities of the specimen and Hg respectively; i_s, i_{Hg} and i_0 the dynamometer restoring currents for the tube containing the specimen of mass m_s, mercury of mass m_{Hg}, and vacuum, respectively, in the same field; then

$$\chi_s/\chi_{Hg} = (m_{Hg}/m_s)[(i_s^2 - i_0^2)/(i_{Hg}^2 - i_0^2)] = (m_{Hg}/m_s)[I_s^2/I_{Hg}^2].$$

A graph of I_U^2 ($I_U^2 = I_s^2$, but, strictly speaking, contains a small correction for twist in the suspension), plotted against I_{Hg}^2 is shown in Figure 1, curve (*a*).

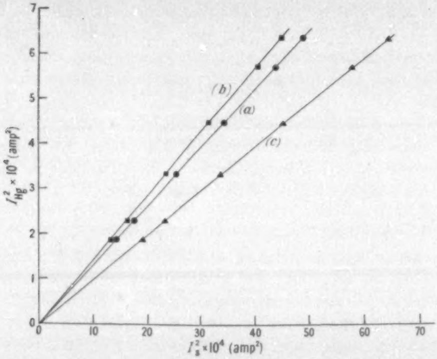


Figure 1. Test for absence of ferromagnetism.

It is undoubtedly linear, so that both the couple on the uranium specimen and that on the mercury were strictly proportional to $H(dH/dr)$. Consequently, the uranium could not have contained a ferromagnetic impurity acting as such. It is easy to provide a proof that ferromagnetic impurities in both the specimen and mercury could not give the line (*a*) of Figure 1.

Applying the Gaussian theory of errors to line (*a*) the value of the slope I_{Hg}^2/I_s^2 was found, whence the final result for the susceptibility of pure uranium was $1{\cdot}740 \pm 0{\cdot}003 \times 10^{-6}$ E.M.U. per gm. The error $\pm 0{\cdot}003$ is the probable error to be attached to the relative measurement; the maximum error is thus $\pm 0{\cdot}015$ and the standard deviation $\pm 0{\cdot}005$.

After these measurements had been completed the specimen was analysed and found to contain 700 parts in 10^6 of iron, 60 in 10^6 of nickel and 5 in 10^6 of cobalt. Consequently, uranium behaves like aluminium and copper in that it may dissolve iron in such a way that the latter cannot exhibit ferromagnetism. The presence of so much iron in a ferromagnetic state would have given couples some twenty times as large as those observed above.

In view of this unexpected behaviour, steps were taken to acquire specimens of nominal iron contents of 5 and 10 atomic per cent. The 5% specimen was too small to fit exactly into the specimen tube so it was fitted with a cylindrical sheath of 'Perspex', for which due allowance was made in the measurements. The results are shown in curve (*b*) of Figure 1; curve (*c*) gives the results for the 10% specimen. Using the same theoretical procedure as before, the susceptibilities of the specimens were found to be $+1{\cdot}842 \pm 0{\cdot}003$ and $1{\cdot}924 \pm 0{\cdot}004 \times 10^{-6}$ E.M.U. per gm. respectively; and, even with as much as 10 atomic per cent impurity of iron, there was still no trace of ferromagnetism.

The variations of the susceptibilities of the three materials with temperature, measured as described, are shown in Figure 2. In some of these measurements convection currents caused some small displacement of the zero of the suspended system. Allowance for this was adequate at lower temperatures, but appeared uncertain above 350° c., and in view of the relatively small changes of susceptibility with temperature, measurements were not recorded above 350° c.

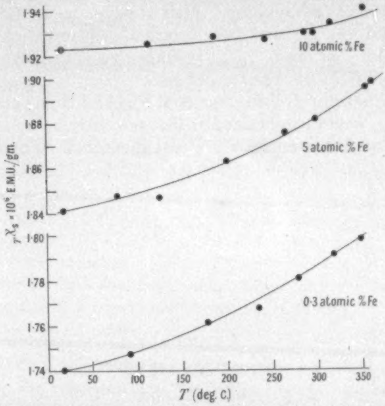


Figure 2. Variation of susceptibility of uranium–iron specimens with temperature.

§ 3. DISCUSSION OF RESULTS

It is understood that uranium forms alloys $U_6.Fe$ and $U.Fe_2$ with iron. In alloys of low iron content nearly all the iron would be expected to be in combination as $U_6.Fe$. Assuming no interaction between uranium metal and $U_6.Fe$ the susceptibility of an iron–uranium alloy should be the sum of the susceptibilities of pure uranium and the compound $U_6.Fe$ according to Wiedemann's law; so that, if a specimen contains x atomic per cent of iron, there will be present a mass $(xM_{Fe}+6xM_U)$ of $U_6.Fe$, where M_{Fe} and M_U are the atomic weights of iron and uranium respectively. Hence, the susceptibility χ_s of such a preparation should be given by

$$\chi_s[xM_{Fe}+(100-x)M_U]=\chi_{U_6.Fe}(xM_{Fe}+6xM_U)+\chi_U(100-7x)M_U. \qquad \ldots\ldots(1)$$

Now the correct iron contents for the three materials as established by later analysis were 0·2978, 4·884 and 9·491 atomic per cent, respectively. Consequently, we have the three equations:

$$23{,}760(1{\cdot}740 \pm 0{\cdot}003)10^{-6} = 442{\cdot}1\chi_{\mathrm{U_6.Fe}} + 23{,}320\chi_{\mathrm{U}}, \qquad \ldots\ldots(2)$$

$$23{,}930(1{\cdot}842 \pm 0{\cdot}003)10^{-6} = 7252\chi_{\mathrm{U_6.Fe}} + 15{,}670\chi_{\mathrm{U}}, \qquad \ldots\ldots(3)$$

$$22{,}080(1{\cdot}924 \pm 0{\cdot}004)10^{-6} = 14{,}090\chi_{\mathrm{U_6.Fe}} + 7{,}991\chi_{\mathrm{U}}. \qquad \ldots\ldots(4)$$

These three equations containing two unknowns will not give the best values of χ_{U} and $\chi_{\mathrm{U_6.Fe}}$ if taken in pairs and treated simultaneously, as each result so obtained does not take account of all the experimental data. As there are only three equations, we used, not the method of least squares, but the more simple 'least sum' method of computation. Half of equation (3) was added to each of equations (2) and (4) to give two 'normal' equations which were solved simultaneously for χ_{U} and $\chi_{\mathrm{U_6.Fe}}$, giving $\chi_{\mathrm{U}} = +1{\cdot}740 \pm 0{\cdot}003 \times 10^{-6}$ and $\chi_{\mathrm{U_6.Fe}} = +2{\cdot}036 \pm 0{\cdot}007 \times 10^{-6}$ E.M.U. per gm.

By taking values from all three curves of Figure 2 at any chosen temperature, then χ_{U} and $\chi_{\mathrm{U_6.Fe}}$ could be obtained at that temperature by calculations exactly similar to those for room temperature just described. Figure 3 (*a*) shows the curve obtained for uranium metal, and it shows that there is a steady increase of susceptibility with temperature of the order of 3·5 per cent over the range 20–350° c. The corresponding curve for $\chi_{\mathrm{U_6.Fe}}$ is given in Figure 3 (*b*), which apparently shows a very small temperature decrease of $\chi_{\mathrm{U_6.Fe}}$ over the same range. The arrows at 20° c. indicate the probable error in this part of the work, viz. ±0·007, whilst those at 200° c. show the probable error in any of the other temperature measurements, viz. ±0·008. This error is larger, because of the errors to be associated with the high temperature value relative to the room temperature value. It is seen that the error range indicated shows that no great reliance can be placed on the small decrease of $\chi_{\mathrm{U_6.Fe}}$ with temperature, i.e. within probable experimental error $\chi_{\mathrm{U_6.Fe}}$ is independent of temperature in the range 20–350° c.

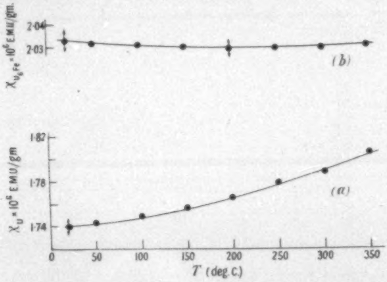


Figure 3. **Variation** of susceptibilities of U and U$_6$.Fe with temperature.

It must be pointed out that these results give no real evidence to support the assumption that the iron actually exists as U$_6$.Fe. Now, the equation (1) could be modified because each specimen might be supposed to contain, for example, uranium and some non-ferromagnetic form of iron not combined with uranium as a compound, or even to contain uranium and UFe_2, which is also known to exist. It was therefore thought that it might be possible, by considering the final error to be attached to χ_U as derived from these various forms of equation (1), to obtain an indication as to which form of equation (1) is the most correct. Calculations were carried out using the relation

$$\chi_s(xM_{Fe}+(100-x)M_U)=\chi_{Fe}xM_{Fe}+\chi_U(100-x)M_U,$$

i.e. assuming only uranium and iron in a paramagnetic form to be present. The value obtained for χ_U was very close to the one quoted above, and the error to be attached to the value was still of the order of $\pm 0{\cdot}003$, thus rendering the calculations inconclusive.

No account has yet been taken of other ferromagnetic elements present as impurity, viz. cobalt and nickel. Cobalt can be completely neglected since all specimens contained less than 20 parts in 10^6. Nickel, however, was present in about 95 parts in 10^6 and, assuming that nickel was present as $U_6.Ni$, equation (1) was written in the form

$$\chi_s[xM_{Fe}+zM_{Ni}+(100-x-z)M_U]$$
$$=\chi_{U_6.Fe}(xM_{Fe}+6xM_U)+\chi_{U_6.Ni}(zM_{Ni}+6zM_U)+\chi_U(100-7x-7z)M_U,$$

where z is the atomic per cent of nickel present. From the three equations so obtained, one for each specimen, values of χ_U, $\chi_{U_6.Fe}$ and $\chi_{U_6.Ni}$ could be obtained, and the calculations showed that, as expected, the values of χ_U and $\chi_{U_6.Fe}$ were not appreciably changed. The error to be attached to $\chi_{U_6.Ni}$ was, however, so large that the result was valueless.

It is possible to make some deductions concerning the behaviour of pure uranium by noting that the experimental results of Figure 3 fit the equation

$$\chi_U=32\times10^{-11}T+1{\cdot}564\times10^{-6}+24{\cdot}0\times10^{-6}/T,$$

since this shows the great preponderance of a temperature independent term in the expression for the susceptibility. If we isolate the last term in the above equation, we find that it is equivalent to a magnetic carrier of $0{\cdot}21\,\mu_B$ obeying the Curie law. It is interesting that Pauling (1947, see also Zachariasen 1949), suggested the allocation to uranium of 5·78 instead of 6 valency electrons, from which we now deduce a possible magneton number of 0·22 for the electron residue; the agreement is somewhat striking.

Suppose, however, that in uranium metal we postulate that the six valency electrons are free and that we allow $-24{\cdot}3\times10^{-6}$ E.M.U. per gm. atom for the radon core, then the gramme atomic paramagnetism at room temperature is $438{\cdot}6\times10^{-6}$. Applying Stoner's (1935, 1936) free electron theory we find that the theoretical susceptibility should be $32{\cdot}1q/V_0\times10^{-6}$ where q is the number of free electrons per atom and V_0 is the width of the energy band in electron volts. The experimental data give $V_0=0{\cdot}4392$ compared with the value 15·52 for completely free electrons. Allowing for the constant term in the preceding paragraph the value would be approximately 4·6. The calculated, $(\Delta\chi_U/\chi_U\cdot\Delta T)$, temperature decrease of susceptibility, assuming that the characteristic temperature of uranium is 4220° K., is $-0{\cdot}29\times10^{-4}$ per degree C.; this is opposite in sign to the observed effect, but the discrepancy between experiment and theory may to some extent be attributed to the thermal expansion of the metal. Indeed, the values of the coefficient of linear expansion, namely 16×10^{-6} and 23×10^{-6} per degree C. in the temperature ranges 20 to 100° C. and 300 to 400° C., respectively, would lead us to expect a small increase of susceptibility with temperature, although it would certainly be much smaller than the experimental value.

It is not thought necessary to compare our results with those of Owen (1912) and Honda (1910) because these workers are unlikely to have used reasonably pure uranium. Since this work was completed the paper by Bommer (1941) has come to hand. He prepared a mixture of uranium and potassium chloride by heating UCl_4 with excess metallic potassium, later distilling off the excess potassium. Bommer's value of χ_U at room temperature is 620×10^{-6} E.M.U. per gm. atom, and he finds it to decrease slightly with increase in temperature. However, as one of his specimens showed ferromagnetism at liquid air temperatures, while two others did not, we feel that there is no point in discussing his results in detail. To carry out such low temperature measurements a new apparatus is necessary, and this is now being made.

Perhaps a few words are pertinent here concerning the evidence for the existence of 5 f electrons in uranium; much of it has recently been summarized by Meggers (1947). We have Mayer's (1941) theoretical treatment of the potential functions of 4 f and 5 f electrons, from which he concluded that a series of 'transuranic elements' should commence filling 5 f shells at $Z=90$ to 92. We have the chemical evidence of Seaborg (1946), Emeléus (1947) and Starke (1947) indicating that the first 5 f electron appears at $Z=90$. Then there is the spectroscopic evidence of Kiess, Humphreys and Laun (1946) who conclude from the spectrum of uranium that three 5 f electrons are present. The heavier metallic elements which show temperature increase of paramagnetism are Ru, Rh and Ir, belonging to the palladium and platinum groups, and it may be significant that in all cases the ground term is usually considered as an F state.

To sum up, the magnitude of the susceptibility and the absence of a Curie or Curie–Weiss temperature relation point to the conclusion that uranium must be treated as a transition element.

ACKNOWLEDGMENTS

We record our thanks to the Director and to Mr. B. W. Mott of the Atomic Energy Research Establishment, Harwell, for the provision of samples of uranium and uranium–iron alloys and for their interest in this work. One of us (J. R. M.) thanks the Department of Scientific and Industrial Research for a maintenance allowance.

REFERENCES

Bates, L. F., 1948, *Modern Magnetism* (Cambridge : University Press).
Bates, L. F., and Somekh, E. M., 1944, *Proc. Phys. Soc.*, **56**, 182.
Bommer, H., 1941, *Z. Anorg. u. Allgem. Chemie*, **247**, 249.
Emeléus, H. J., 1947, *Some Aspects of Nuclear Chemistry* (The Royal Institute of Chemistry).
Honda, K., 1910, *Ann. Phys., Lpz.*, **32**, 1044.
Kiess, C. C., Humphreys, C. J., and Laun, D. D., 1946, *J. Res. Nat. Bur. Stand., Wash.*, **37**, 57.
Mayer, M. G., 1941, *Phys. Rev.*, **60**, 184.
Meggers, W. F., 1947, *Science*, **105**, 514.
Owen, M., 1912, *Ann. Phys., Lpz.*, **37**, 657.
Pauling, L. C., 1947, *J. Amer. Chem. Soc.*, **69**, 542.
Seaborg, G. T., 1946, *Science*, **104**, 379.
Starke, K., 1947, *Naturwissenschaften*, **69**, 34.
Stoner, E. C., 1935, *Proc. Roy. Soc.* A, **152**, 672; 1936, *Ibid.*, **154**, 656.
Zachariasen, W. H., 1949, *Acta Crystallogr.*, **2**, 94.

On Phase-Contrast Testing with a Slit Source

By E. H. LINFOOT
The Observatory, Cambridge

MS. received 10*th November* 1949

ABSTRACT. In the standard form of the phase-contrast test, the light-source is a pinhole comparable in size with the Airy disc of the system under test. As a consequence, it is difficult to obtain enough light through the pinhole to carry out the test with full efficiency. If a slit source and a phase-changing strip are used this difficulty is overcome, but the properties of the test are altered; for example, the brightness distribution seen on the surface of a true mirror under the test is no longer radially symmetrical.

In the present paper the properties of the phase-strip test are examined in the mathematically simplest case where the light-source is a slit of negligible width, passing monochromatic light, and it is concluded that the test can safely be used in figuring an optical surface to an accuracy of about one twentieth of a fringe. Below this limit, the systematic error inherent in the usual interpretation of the test begins to be appreciable in the case of slow errors, though rapid local errors are correctly seen right up to the limit of sensitivity of the test.

§ 1. INTRODUCTION

The elegant and ingenious phase-contrast method of testing the figures of optical surfaces, devised about sixteen years ago by F. Zernike (1934 a), has been described from the experimental side by C. R. Burch (1934) and from the theoretical by its originator and by the present writer (Zernike 1934 b, Linfoot 1946 a). Dr. Burch has described the setting up of the test essentially as follows:

An artificial star of small diameter is placed at one of the stigmatic points of an optical system and at the other stigmatic point there is placed the small circular phase-retarding disc or Zernike disc. The star and the Zernike disc are both preferably somewhat smaller than the Airy disc of the system at the points at which they are respectively placed. The eye of the observer is placed close behind the Zernike disc, and he looks either directly at the surface under test or at its image in any auxiliary surfaces which may come between it and him. He then sees interference fringes formed between the wave leaving the surface and the supplementary wave to which its passage through the Zernike disc gives rise. If the Zernike disc is not larger than the Airy disc produced by the system, that part of the supplementary wave lying within the cone of light proceeding from the system will be substantially spherical, so that (since the wave front leaving the system departs from sphericity only by the errors of the mirror under test) the fringe system seen in a given wavelength will be substantially the Newton-ring system which would be seen if a true-figure test plate were laid on, or very close to, the surface under test. The colour sequence of the fringes will be determined by the chromatic difference of phase-retardation of the Zernike disc, and can be ascertained from the colour changes which occur when the disc is moved out of focus, and when it is decentred.

A practical drawback to this form of the test is the difficulty of getting enough light through the very small pinhole which has to be used. A high pressure mercury arc used with a good-quality condenser will overcome the difficulty to some extent, but even so the test needs to be carried out in a darkened room, and in testing an unsilvered glass mirror the amount of light is inconveniently small.

In the Foucault knife-edge test, this shortage of light can be overcome very simply by substituting a fine slit for the pinhole and testing with the knife-edge set as accurately parallel as possible to the image of the slit. The forms and relative intensities of the shadows seen under the test are not altered by the change, provided only that the slit is too short for the aberrations of the system to vary appreciably along its length, and the amount of light available for testing purposes is increased by a factor of 50 or 100.

It was pointed out by Zernike that a slit-source can also be used in the phase-contrast test provided that the phase-changing disc is replaced by a phase-changing strip, adjusted to cover the diffraction image of the slit. Strips of suitable dimensions can be made without very much difficulty by pressing fine threads of resin between glass, or by stroking the surface of a glass plate with a 'cutanit' (molybdenum-titanium carbide) tool*, and fairly rapid irregular errors are found to show very nearly the same fringe-patterns under test with a strip as with a disc.

Thus the phase-strip test shares one of the most valuable properties of the Zernike-disc test, that of displaying rapid irregular errors of figure on the mirror surface in a way which allows their size and shape to be correctly 'seen' by simple inspection of the surface under test.

Of even greater importance to the practical mirror-maker, however, is the question whether the strip test can safely be used for the control of the three errors which are most commonly encountered during the figuring of an optical surface, namely primary spherical aberration, zonal error and astigmatism. The examination of this question is the main object of the present paper.

In §2 formulae are derived which predict the intensities seen under the test on a surface with arbitrary smooth errors. In §3, these formulae are applied to the case of a nearly true mirror and it is shown that something near to the best results may be expected with a quarter-wave strip, of width about one third of the diameter of the Airy disc. The appearance, under test with such a strip, of small amounts of primary astigmatism, of primary spherical aberration, and of local zonal error is discussed in §4, and from the examination of special cases it is inferred that the test can safely be used for the figuring of optical surfaces to an accuracy of about one twentieth of a fringe.

Throughout the analysis it is assumed that the absorption and back-reflection of the strip may be disregarded, that the width of the slit is small compared with the Airy disc of the system under test and that the light is monochromatic. In practice, the width of the slit would usually be comparable in size with the diameter of the Airy disc and the light polychromatic, but a discussion of the extent to which the properties of the test are modified by these circumstances is beyond the scope of the present paper.

§2. DIFFRACTION THEORY OF THE PHASE-CONTRAST STRIP TEST

We consider first the case where the light originates in a point source, the geometrical image of which falls on the centre of the phase-changing strip. The notation and assumptions are similar to those used in earlier papers (Linfoot 1946 a, b). For convenience they are given again here.

---

* C. R. Burch (1934). The engraved strips advance the phase, giving 'positive phase contrast'.

All the cases wher a convergent pencil is being null-tested can be covered by supposing that the wave fronts originate at the surface of a nearly spherical mirror M which is being tested at its approximate centre of curvature.. The results of the test are then interpreted in terms of errors of figure of the mirror M. We suppose that these errors of figure may amount to several wavelengths, but that the error slopes on the mirror are not so steep as to spread out the visible image to more than a moderate multiple, say 5 or 10, of the size of the Airy disc.

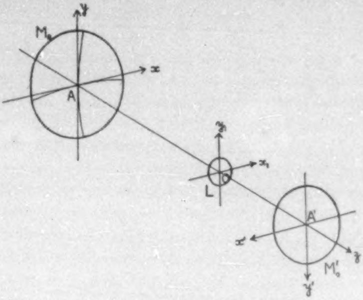


Figure 1.

Then it makes no appreciable difference if we suppose the wave to be leaving a true spherical surface M_0 (Figure 1), lying everywhere within a few wavelengths of the surface M, and suppose that at the point (x, y, z) on M_0 the complex displacement is

$$E(x,y) = |E(x,y)| \exp\{(2\pi i/\lambda)\phi(x,y)\};$$

thus $|E(x,y)|$ is the amplitude at the point (x,y) on M, and $\phi(x,y)$ the phase there, while λ is the wavelength of the light. The variation in $\phi(x,y)$ expresses the distortion of the wave fronts. Outside the boundary-circle C, lying in M, and corresponding to the edge of the mirror, $E(x,y)$ is defined as zero.

After leaving the mirror, the light comes to a more or less imperfect focus in the neighbourhood of the centre of curvature O of M, at a distance s from its pole A; we suppose that s is large compared with the diameter of the mirror. Just behind O, a viewing system L receives the light and forms and image M_0' of the surface M; we denote by C′ the boundary of M_0'.

At O is placed the phase-changing strip which renders variations of phase in the wave fronts leaving M_0 visible as variations of intensity in the image-surface M_0'; x_1 and y_1, measured parallel to x and y, are coordinates in the space near O. The central line of the strip lies along the axis Oy_1 in the plane T drawn through O at right angles to Oz.

We define a system of coordinate-numbers (x', y') in the surface M_0', by assigning to the image P′ of the point P (x, y, z) on M_0 the coordinate-numbers $x' = x, y' = y$. The intensity at P′ is thus the same thing as the 'intensity seen under the test' at P. To avoid unessential complications, we suppose that L is a perfect optical system.

By the usual arguments, based on an application of Huyghens' principle, the complex displacement in the intermediate image surface, which we take to be the sphere S of centre A passing through O, is

$$W(u,v)=\frac{1}{2\pi}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}e^{ixu+iyv}E(x,y)\,dx\,dy, \qquad \ldots\ldots(2.1)$$

where

$$u=\frac{2\pi x_1}{\lambda s},\quad v=\frac{2\pi y_1}{\lambda s}, \qquad \ldots\ldots(2.2)$$

and the integral is only formally over an infinite domain, since the integrand vanishes outside the boundary contour C.

From the assumptions made concerning the form of the errors of the surface under test, it follows that almost all the light crosses the intermediate image surface in the immediate neighbourhood of the point O.

If the phase-changing strip were not present, the complex displacement $D(x',y')$ at the point (x',y') in the final image-surface would be obtained by integrating the product

$$\frac{1}{2\pi}\exp\{-i(x'u+y'v)\}\,.\,W(u,v)$$

over the area $u^2+v^2\leqslant R^2$ of the u,v plane which falls within the aperture of the viewing system L. (In the case of an f/10 pencil and a viewing telescope of 1 inch aperture, the value of R is approximately 1,300.) The effect of an idealized Zernike strip occupying the region $-c\leqslant u'\leqslant c$ of the intermediate image surface and retarding by α radians the phase of the waves passing through it can be represented by changing $W(u,v)$ into $e^{-i\alpha}W(u,v)$ throughout this region and leaving it unaltered elsewhere. In saying this, we assume that the effect of the strip is merely to change the phase of those parts of the wave which impinge upon it, without influencing the neighbouring parts. This assumption is never strictly true, but it is a permissible approximation when the pencil under test is of small numerical aperture.

In reality, the phase-strip is not an infinitesimally thin lamina but a solid body of length $10{,}000\lambda$ or more, of width 10 to 100λ, and of thickness comparable with λ. Now throughout that u,v region of S in which an alteration in the values of $W(u,v)$ can cause a visible change in the appearance of the mirror under the test, the distance between the surface S and the plane T is of the order of $10^{-4}\lambda$ in a practical case. For example, when a 12-inch mirror is tested at f/10 in light of wavelength $\lambda=2\times10^{-5}$ inch, the distance between S and T is less than $5\times10^{-5}\lambda$ throughout the circle $u^2+v^2\leqslant14{\cdot}7\times10^4$ of radius 100 times that of the Airy disc, while the replacement of $W(u,v)$ by zero outside this circle does not sensibly affect the appearances seen under the test.* It follows that, to the order of approximation here involved, the idealized phase-changing lamina which replaces our three-dimensional phase-strip in the analysis may as well be supposed to lie in the intermediate image surface S as in the surface T. The complex displacement at the point (x', y') of the final image surface is therefore given, to a sufficient approximation, by the equation

---

* This may be verified experimentally by testing a nearly true mirror of long focal ratio (about f/50) with an iris diaphragm placed in front of the viewing telescope. (See Linfoot (1946 b) for an account of the corresponding experiment on changes in the appearances seen under the Foucault knife-edge test.)

$$D(x', y') = \frac{1}{2\pi} \iint_{u^2+v^2 \leqslant R^2} \exp(-iux' - ivy') \, . \, W(u, v) \, du \, dv$$
$$+ \frac{1}{2\pi} (e^{-i\alpha} - 1) \iint_{u^2+v^2 \leqslant R^2, \, -c \leqslant u \leqslant c} \exp(-iux' - ivy') \, . \, W(u, v) \, du \, dv. \qquad \ldots\ldots(2.3)$$

The first term of (2.3) is simply the complex displacement which results when the mirror is imaged through the viewing system, without the interposition of the Zernike strip, and the second term therefore describes the effect of the strip.

We next replace R by ∞ in (2.3). This amounts to neglecting the effects of the finite aperture of the viewing telescope. For mirrors with errors of the type under consideration, the resulting change in the predicted intensities is too small to be detected by the eye, except at points whose distance from the rim of the mirror is comparable with the resolving power of the viewing system. If, in order to cover the possibility of a central obstruction or of a sharp step on the surface of M, the hypotheses are widened by allowing $E(x, y)$ to be discontinuous along a finite set of arcs of differentiable curves, the above assertion remains valid provided we replace the words "rim of the mirror" by "set of points of discontinuity of $E(x, y)$". At these points the intensities depend critically on the aperture and optical aberrations of the viewing system, as well as on the errors of the mirror, and we therefore exclude them from the discussion. At the remaining points we obtain, on setting $R = \infty$ in (2.3), the equation

$$D(x', y') = \frac{1}{2\pi} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \exp(-iux' - ivy') \, . \, W(u, v) \, du \, dv$$
$$+ \frac{1}{2\pi} (e^{-i\alpha} - 1) \iint_{-c \leqslant u \leqslant c, \, -\infty < v < \infty} \exp(-iux' - ivy') \, . \, W(u, v) \, du \, dv. \qquad \ldots\ldots(2.4)$$

On substituting for $W(u, v)$ from (2.1) and applying Fourier's inversion formula, (2.4) becomes

$$D(x', y') = E(x', y') + \frac{1}{\pi} (e^{-i\alpha} - 1) \int_{-c}^{c} e^{-iux'} \, du \int_{-\infty}^{\infty} e^{iux} E(x, y') \, dx$$
$$= E(x', y') + \frac{1}{\pi} (e^{-i\alpha} - 1) \int_{-\infty}^{\infty} E(x, y') \frac{\sin c(x - x')}{x - x'} \, dx. \qquad \ldots\ldots(2.5)$$

These transformations are justified because $E(x, y)$ is zero everywhere outside a finite region of the x, y plane, is of bounded variation in this region, and is differentiable at the point $x = x'$, $y = y'$.

The intensity seen under the test, at the point (x', y') on the mirror-surface or in the halo, is then measured by the quantity

$$I(x', y') = |D(x,' y')|^2. \qquad \ldots\ldots(2.6)$$

So far, the light-source has been assumed to be a pinhole of negligible dimensions. In the case of a slit of negligible width, each point of the slit can be regarded as a point source. If each point of the primary light source were sharply focused on to the plane of the slit by a wide-aperture condenser system of negligible aberrations, the different points of the slit could be treated as independent sources. In such a case, the observed intensity under the test would simply be the sum of the intensity contributions from the different point sources. If we suppose that the length of the slit is not great enough for off-axis aberrations of the system to make an appreciable contribution to the apparent errors of the surface under test, the intensity distributions contributed by any two points of the slit differ only by a constant factor. Therefore, the intensity function $I(x',y')$ is still given by equation (2.6).

However, in a practical case the aperture of the condenser system is often comparable with that of the system under test and it is not true that all points of the primary source are sharply focused on to the plane of the slit by the condenser. Each luminous particle W of the primary source contributes a wave displacement

$$E_{\mathrm{W}}(x,y)=A_{\mathrm{W}}(x,y)E(x,y)$$

on the surface of M_0, where $E(x,y)=\exp\{(2\pi i/\lambda)\phi(x,y)\}$ inside C and zero outside C, as before, while $A_{\mathrm{W}}(x,y)$ expresses the characteristics of the wave emitted by the W particle when, after passing through the condensing system and the slit, it arrives at the point (x,y) on the nearly spherical mirror M. Because the width of the slit is negligible, $A_{\mathrm{W}}(x,y)$ does not vary appreciably with x, and we may write, to a sufficient approximation

$$E_{\mathrm{W}}(x,y)=A_{\mathrm{W}}(y)E(x,y). \qquad \ldots\ldots(2.7)$$

$|E_{\mathrm{W}}(x,y)|^2$ measures the intensity of the wave at this point, so that if the mirror is to appear uniformly lit in the absence of the phase strip

$$\sum_{\mathrm{W}}|E_{\mathrm{W}}(x,y)|^2$$

must be constant over its surface. By a suitable choice of units we can write this condition in the form

$$\sum_{\mathrm{W}}|A_{\mathrm{W}}(y)|^2=1 \quad \text{on } \mathrm{M}_0. \qquad \ldots\ldots(2.8)$$

The complex displacement contribution from the W particle in the final image surface is then obtained on replacing $E(x,y)$ by $A_{\mathrm{W}}(y)E(x,y)$ in the previous analysis; it is

$$\begin{aligned} D_{\mathrm{W}}(x',y')&=A_{\mathrm{W}}(y')E(x',y')+\frac{1}{2\pi}(e^{-i\alpha}-1)\int_{-\infty}^{\infty}A_{\mathrm{W}}(y')E(x,y')\,\frac{\sin c(x-x')}{x-x'}\,dx \\ &=A_{\mathrm{W}}(y')D(x',y'), \end{aligned}$$

where $D(x',y')$ is given by (2.5). Since the waves from the different W particles are mutally incoherent, the total intensity seen on M_0 under the test is now

$$\begin{aligned} \sum_{\mathrm{W}}|D_{\mathrm{W}}(x',y')|^2&=\sum_{\mathrm{W}}|A_{\mathrm{W}}(y')|^2|D(x',y')|^2 \\ &=|D(x',y')|^2, \end{aligned}$$

by (2.8). That is, the intensities seen on M_0 are still given by (2.6). The same is true of the intensities in the halo, since these differ from zero only in the y-range where (2.8) holds.

From (2.5) we see that the value of $D(x',y')$ at any point (x',y') depends only on the figure of the mirror M along the horizontal line $y=y'$ and that the halo is confined to a horizontal band filling the same y-range as do the points of the mirror. In these two respects the phase-strip test resembles the knife edge test. On the other hand, since the function $\sin\{c(x-x')\}/(x-x')$ is equivalent to a function of x everywhere continuous, it can be shown that there is no bright 'Rayleigh ring' at the edge of the mirror, but an apparent ordinary discontinuity in $I(x',y')$ arising from the first term on the right of (2.5). The same is true at a step-discontinuity on the surface of M.

§ 3. NEARLY TRUE MIRROR; CHOICE OF PHASE STRIP

On a nearly true mirror of uniform reflecting power, we can write

$$\begin{aligned} E(x,y) &= 1+iE^*(x,y) \qquad (x^2+y^2 \leqslant H^2) \\ &= 0 \qquad\qquad\qquad\quad (x^2+y^2>H^2), \end{aligned} \qquad \ldots\ldots(3.1)$$

where $E^*(x,y)$ is real and so small that its square can be neglected in comparison with unity, while $2H$ is the diameter of the mirror surface.* (2.5) then gives, for $x^2+y^2 \leqslant H^2$,

$$\begin{aligned} D(x,y) &= 1+iE^*(x,y)-[(1-\cos\alpha)+i\sin\alpha]\frac{1}{\pi}\int_{-Y}^{Y}(1+iE^*(t,y))\frac{\sin c(t-x)}{t-x}\,dt, \\ &= 1-(1-\cos\alpha)I_c(Y)+\sin\alpha\,\frac{1}{\pi}\int_{-Y}^{Y}E^*(t,y)\frac{\sin c(t-x)}{t-x}\,dt \\ &\quad +i\left[E^*(x,y)-\sin\alpha\,I_c(Y)-(1-\cos\alpha)\frac{1}{\pi}\int_{-Y}^{Y}E^*(t,y)\frac{\sin c(t-x)}{t-x}\,dt\right], \end{aligned}$$

$$\ldots\ldots(3.2)$$

where Y is written for $(H^2-x^2)^{1/2}$ and

$$I_c(Y)=\frac{1}{\pi}\int_{-Y}^{Y}\frac{\sin c(t-x)}{t-x}\,dt=\frac{1}{\pi}\,[\mathrm{Si}\{c(Y-x)\}+\mathrm{Si}\{c(Y+x)\}]. \qquad \ldots\ldots(3.3)$$

Hence

$$\begin{aligned} I(x,y) &= |D(x,y)|^2 = 1-4\sin^2\frac{\alpha}{2}\,I_c(Y)[1-I_c(Y)] \\ &\quad -2\sin\alpha\left[I_c(Y)E^*(x,y)-\frac{1}{\pi}\int_{-Y}^{Y}E^*(t,y)\frac{\sin c(t-x)}{t-x}\,dt\right] \end{aligned}$$

$$\ldots\ldots(3.4)$$

$$= I_0(x,y)-2I_c(Y)\sin\alpha E^*(x,y)+\frac{2}{\pi}\sin\alpha\int_{-Y}^{Y}E^*(t,y)\frac{\sin c(t-x)}{t-x}\,dt,$$

$$\ldots\ldots(3.5)$$

where

$$I_0(x,y)=1-4\sin^2\frac{\alpha}{2}\,I_c(Y)[1-I_c(Y)] \qquad \ldots\ldots(3.6)$$

is the intensity distribution on a true mirror.

---

* Thus a mirror whose errors of figure nowhere exceed 1/20 fringe may be regarded as nearly true in the above sense.

For $|x| > Y$, $|y| \leqslant H$, that is to say in the halo, we have in place of (3.2) and (3.5)

$$D(x,y) = -[(1-\cos\alpha) + i\sin\alpha]\,\frac{1}{\pi}\int_{-Y}^{Y}(1+iE^*(t,y))\,\frac{\sin c(t-x)}{t-x}\,dt,$$

$$I(x,y) = |D(x,y)|^2 = [(1-\cos\alpha)^2 + \sin^2\alpha]$$

$$\times\left[(I_c(Y))^2 + \left(\frac{1}{\pi}\int_{-Y}^{Y} E^*(t,y)\,\frac{\sin c(t-x)}{t-x}\,dt\right)^2\right]$$

$$= 4\sin^2\frac{\alpha}{2}\,(I_c(Y))^2$$

$$= I_0(x,y). \qquad \ldots\ldots(3.7)$$

Thus the halo is unaffected by small errors of figure on the mirror.

A small local error on an otherwise true mirror corresponds to the case where $E^*(x,y)$ vanishes everywhere outside a small x interval J of length $\eta \ll H$. For each fixed y, the integral on the right of (3.5) is then $O(\eta)$ times the average value of $E^*(x,y)$ in the interval J and the intensity distribution is therefore represented to a good approximation by the equation

$$I(x,y) = I_0(x,y) - 2I_c(Y)\sin\alpha\,.\,E^*(x,y) \qquad \ldots\ldots(3.8)$$

$$= 1 - 4\sin^2\frac{\alpha}{2}\,I_c(Y)[1-I_c(Y)] - 2I_c(Y)\sin\alpha\,.\,E^*(x,y). \qquad \ldots\ldots(3.9)$$

Equation (3.9) shows that on a small local bump the intensity is reduced by an amount very nearly proportional to $\sin\alpha$ times the height of the bump. In the neighbourhood of such a bump the 'background intensity' $I_0(x,y)$ is nearly constant, by (3.6) and (3.3). It follows that here, as in the Zernike disc test, the observed intensity change is, within the limits of visual estimation, proportional to the height changes over the bump, and the dark patch which (with a phase-retarding strip) represents the error corresponds to it in shape and intensity. What is even more important from the practical point of view, its boundary agrees exactly with the boundary of the bump, so that the mirror-maker does not find (as he does with the Foucault test) that his polisher has inexplicably trespassed outside the limits of the bump.* As a result, local figuring operations can be carried out with ease and certainty under the phase-strip test which would be impracticable under the Foucault test.

In the case where shallow irregular error covers a large part of the mirror surface, the approximation (3.9) is no longer applicable, and the more accurate formula (3.4) must be used. But since the value of the integral

$$\rho(x,y) = \frac{1}{\pi}\int_{-Y}^{Y} E^*(t,y)\,\frac{\sin c(t-x)}{t-x}\,dt \qquad \ldots\ldots(3.10)$$

changes only slowly as x and y vary, except near the points $(0, \pm H)$, the intensity on and near a rapid local error is still given by an equation

$$I(x,y) = -2I_c(Y)\sin\alpha E^*(x,y) + \text{background intensity}, \qquad \ldots\ldots(3.11)$$

and the above conclusions remain valid, except near the top and bottom of the mirror disc.

---

* See Ponomarev, *Bull. de l'Inst. Astronomique*, No. 30, Leningrad 1931, and Linfoot (1948).

Conclusions about the sensitiveness of the test can be drawn from (3.5) and (3.7). By (3.7), shallow errors do not appreciably affect the halo; it follows that they do not appreciably alter the total amount of light seen under test on the mirror disc, but merely redistribute it. That is to say

$$\iint_{x^2+y^2\leqslant H^2} I(x,y)\,dx\,dy = \iint_{x^2+y^2\leqslant H^2} I_0(x,y)\,dx\,dy. \qquad \ldots\ldots(3.12)$$

In the case of a small local bump, the second term on the right of (3.5) describes quantitatively the manner in which light is displaced out of the small area covered by the bump and the third term describes the manner in which the displaced light is redistributed over a horizontal band of the mirror surface. Because it is now smoothly spread over a much greater area than before, the intensity changes due to this displaced light are small in comparison with those on the bump itself.

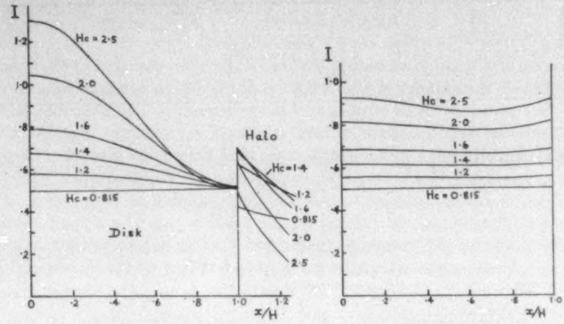
In the case of a 'cobbled surface', a large number of irregular bumps and hollows are distributed at random over the whole region $x^2+y^2\leqslant H^2$. Since the right hand side of (3.5) remains unchanged in value when a constant is added to $E^*(x,y)$, we may suppose that the mean of $I_c(Y)E^*(x,y)$ over the surface is zero. Then it follows from (3.5) and (3.12) that

$$\iint_{x^2+y^2\leqslant H^2} \rho(x,y)\,dx\,dy = 0. \qquad \ldots\ldots(3.13)$$

Since $\rho(x,y)$ varies slowly in comparison with $I_c(Y)E^*(x,y)$, it follows that, except near the points where $E^*(x,y)$ is abnormally small, and near the points $(0, \pm H)$, the third term on the right of (3.5) is small in comparison with the second term. That is, the resultant effect of the redistributed light is small compared with the intensity variations arising from the second term on the right of (3.5). The appearance of the cobbles under the phase-strip test is therefore, to a good approximation, a faithful representation of their true sizes and shapes, except near the top and bottom of the mirror surface.

It follows from (3.4) that the intensity changes by which the errors on a nearly true mirror show themselves under the test are proportional to $\sin\alpha$. The changes are therefore greatest when the phase-strip advances or retards the phase by a quarter-cycle. It cannot, however, be inferred from this that the test is most sensitive when $\alpha = \pm\frac{1}{2}\pi$, since the background intensity $I_0(x,y)$ also depends on α.

If the test is to be sensitive in detecting slow as well as rapid errors, α and c must be chosen so that the intensities seen under the test do not vary too much over the surface of a true mirror. In particular, $I_0(x,y)$ must not vary too much along the main diameter $y=0$, $-H\leqslant x\leqslant H$. Figure 2(*a*) shows the intensities seen along the main diameter and in the halo when a true mirror is tested with $\alpha=\frac{1}{2}\pi$ and with selected values of Hc, the phase-strip being set at focus. In practice, the phase-strip would be set by trial so that the disc appears as nearly as possible uniformly bright. Figure 2(*b*) shows the intensities along the main diameter at the focal settings which would make them as nearly uniform as possible. They were calculated by setting $E^*(x,0)=\epsilon x^2$ in (3.5) and giving ϵ the values, found by trial, $-0{\cdot}065$, $-0{\cdot}13$, $-0{\cdot}21$, $-0{\cdot}41$, $-0{\cdot}70$ in the respective cases $Hc=0{\cdot}815$, $1{\cdot}2$, $1{\cdot}4$, $1{\cdot}6$, $2{\cdot}0$, $2{\cdot}5$. Here c is measured in u, v units, as before, so that $Hc=3{\cdot}83$ in the case of a strip just wide enough to cover the Airy disc completely. In the case $Hc=25$, the above procedure gives a rather rough approximation to the true values of $I(x, 0)$. However, it is already a rather rough approximation to use the focal settings which minimize intensity variations along the main diameter instead of over the whole disc.



(*a*) With strip at focus.

(*b*) With strip adjusted to give greatest uniformity along the main diameter.

Figure 2. Intensities seen along the main diameter of a true mirror tested with quarter-wave phase-retarding strips of different widths.

From Figure 2 (*b*) it appears that, in the case $\alpha = \pm\frac{1}{2}\pi$, Hc should not exceed about 2·0 if a true mirror is to appear uniformly illuminated along the main diameter. The width of the phase-strip is then 0·52 times the diameter of the Airy disc. Greater uniformity along the main diameter, with only a moderate decrease in intensity there, is obtained by choosing $Hc = 1{\cdot}2$. The appearance of the disc with the phase-strip set at focus is then as shown in Figure 3. There is a very noticeable increase in brightness near the top and bottom of the disc, but its appearance compares favourably in uniformity with that under the Foucault test.

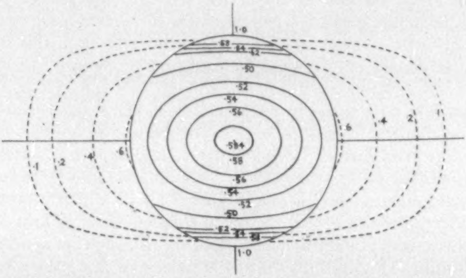


Figure 3. Intensities seen on true mirror under test with quarter-wave phase-retarding strip ($Hc = 1{\cdot}2$, $\alpha = \frac{1}{2}\pi$) set at focus. The dotted isophotes show the intensities in the halo.

A strip with $\alpha = \frac{1}{2}\pi$ and $Hc = 2{\cdot}0$, used at the appropriate focal setting, would not give appreciably worse uniformity over the disc as a whole than one with $Hc = 1{\cdot}2$ and would be about equally useful in practice. However, in the critical examination of weak zonal aberration something is gained by using the narrower strip, together with a slit mask exposing only a horizontal band of the mirror surface around its main diameter. By § 2, the presence of the mask does not alter the intensities on the exposed portion of the mirror surface.

Increasing $|\alpha|$ to a value between $\frac{1}{2}\pi$ and π would, by (3.6), decrease the values of $I_0(x,y)$ and accentuate their fractional variation. Some improvement in the sensitivity of the test for rapid error could be obtained in this way, but at the cost of decreasing its reliability as a means of controlling slow errors.

We therefore select, as a typical special case on which to base the quantitative estimates of sensitivity in the next Section, a strip for which $\alpha = \frac{1}{2}\pi$ and $Hc = 1{\cdot}2$; that is to say a quarter-wave phase-retarding strip of width approximately one third of the diameter of the Airy disc. With such a strip, giving 'negative phase contrast', a local bump appears darker than the neighbouring parts of the surface in accordance with (3.8).

§ 4. SLOW ERRORS; LOCAL ZONAL ERROR

A slow error whose behaviour under the test is of special practical interest is primary astigmatism, obtained on setting

$$\begin{aligned} E(x,y) &= \exp\{i(ax^2+2hxy+by^2)\} \qquad (x^2+y^2 \leqslant H^2) \\ &= 0 \qquad\qquad\qquad\qquad\qquad\quad\;\; (x^2+y^2 > H^2). \qquad \ldots\ldots(4.1) \end{aligned}$$

In testing a slightly astigmatic mirror, the phase-strip is adjusted so that the halo appears symmetrical * and the mirror as evenly and symmetrically illuminated as possible. It results that the strip is set centrally, but that its focal setting depends on the orientation of the mirror and on the amount of the astigmatism.

The effect of moving the phase-strip through a distance δs from O, away from the mirror, can be found by replacing $E(x,y)$ on the right of (2.5) by

$$\exp\{i\eta(x^2+y^2)\} \,.\, E(x,y)$$

and $D(x',y')$ on the left by $\exp\{i\eta(x'^2+y'^2)\}\,.\,D(x',y')$, where $\eta = \pi(\delta s/\lambda s^2)$. The resulting equation gives the new value of $D(x',y')$; on substituting for $E(x,y)$ from (4.1) and simplifying, it becomes

$$D(x',y') = \exp\{i(ax'^2+2hx'y'+by'^2)\}$$
$$\times \left[\genfrac{}{}{0pt}{}{1}{0} - (1-e^{-i\alpha}) \int_{x'-Y}^{x'+Y} \exp\{i(a+\eta)(2x't+t^2)+2ihty'\} \,.\, \frac{\sin ct}{t}\,dt\right], \qquad \ldots\ldots(4.2)$$

where the upper or lower alternative is taken according as $x^2+y^2 \lesseqgtr H^2$.

At the focal setting $\eta = -a$, (4.2) gives

$$I(-x',-y') = I(x',y'); \qquad \ldots\ldots(4.3)$$

at other focal settings this identity does not hold. Thus there is a unique focal setting at which the astigmatic mirror appears skew-symmetrical under the test. This setting is the one which is used in practice; it will be called the preferred setting. (4.2) gives, at the preferred setting

$$I(x',y') = |D(x',y')|^2 = \left|\genfrac{}{}{0pt}{}{1}{0} - (1-e^{-i\alpha})\frac{1}{\pi}\int_{x'-Y}^{x'+Y} \exp 2ihty' \frac{\sin ct}{t}\,dt\right|^2. \qquad \ldots\ldots(4.4)$$

The integral on the right is easily evaluated in terms of the functions Si, Ci.

---

* Compare (3.7) above. If the amount of astigmatism is not small, the halo is skew-symmetrical by (4.3).

Of the coefficients a, h, b in (4.1), only h occurs in the expression (4.4) for the intensity at the preferred setting. It follows that a given astigmatism shows itself most strongly when the mirror is oriented with its astigmatic principal axes at angles of 45° with the direction of the phase-strip. If one of these axes is parallel to the direction of the phase-strip, the appearance of the mirror under the test is the same as that of a true mirror and the astigmatism escapes detection.*

Figure 4 shows the appearance, calculated from (4.4), of $\frac{1}{10}$ fringe of astigmatism, tested with a quarter-wave phase-retarding strip of width approximately one third of the diameter of the Airy disc, the mirror being oriented so that its astigmatic principal axes make angles of 45° with the vertical strip. In this case the astigmatism can easily be recognized under the test, and it may be estimated that one-twentieth of a fringe of astigmatism should be detectable.

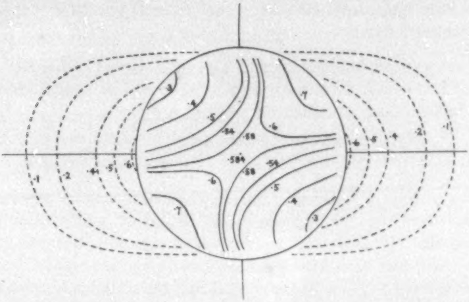


Figure 4. One tenth of a fringe of primary astigmatism ($E(x, y) = \exp\{2\pi ixy/10\}$) under test with a quarter-wave phase retarding strip ($Hc = 1{\cdot}2$, $\alpha = \frac{1}{2}\pi$). The dotted isophotes show the intensities in the halo.

Of equal practical importance is the question whether a small amount of primary spherical aberration can be seen and identified without difficulty. Figure 5 shows the appearance of one tenth of a fringe of primary spherical aberration, of type

$$E(x, y) = \exp\{8\pi i(r^2 - r^4)\}, \qquad r^2 = x^2 + y^2, \qquad \ldots\ldots(4.5)$$

tested with the same strip.† From the figure it appears that a broad dark band, of mean radius approximately $0{\cdot}7H$ and not much differing from a radially symmetrical ring in appearance, indicates correctly the position of the 'crest of the doughnut', while the brightening of the disc towards the centre and edge corresponds sufficiently closely with the variations in the error function $r^2 - r^4$. Thus one tenth of a fringe of spherical over-correction presents a striking and easily recognized picture under the test, and it seems safe to say that one twentieth of a fringe could be identified without difficulty.

---

* Compare the properties of astigmatism under the Foucault test, *Mon. Not. R. Astr. Soc.*, 1945, **105**, 193.

† Calculated from (2.5). My thanks are due to the Cambridge Mathematical Laboratory for their help in computing this figure.

Finally, in order to form an idea of the sensitiveness of the test in detecting local zonal error, we consider an 'artificial' zone, one twentieth of a fringe high all over its surface, occupying the annulus

$$0{\cdot}6H \leqslant r \leqslant 0{\cdot}7H \qquad \ldots\ldots(4.6)$$

of the mirror surface. Evidently $E(x,y)$ has step-discontinuities on its inner and outer boundaries of the zone, as well as at the edge of the mirror, and the immediate neighbourhood of these points, as well as of the rim of the mirror, must therefore be excluded from the discussion. At the remaining points, (2.5) may be applied with

$$\begin{aligned} E(x,y) &= 1 & & 0 \leqslant r \leqslant 0{\cdot}6H \\ &= e^{2\pi i/20} & & 0{\cdot}6H < r < 0{\cdot}7H \\ &= 1 & & 0{\cdot}7H \leqslant r \leqslant H \\ &= 0 & & r > H, \qquad \ldots\ldots(4.7) \end{aligned}$$

where, as before, $r^2 = x^2 + y^2$. The resulting intensity distribution is represented in Figure 6, which shows that the zone will appear as a clearly marked dark ring of about half the brightness of the nearby portions of the mirror surface.

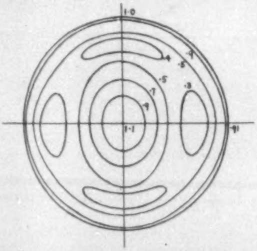


Figure 5. One-tenth of a fringe of spherical over-correction $\left(E(x,y) = \exp\left\{\frac{2\pi i}{10}(r^2 - r^4)\right\}\right)$ under test with a quarter-wave phase-retarding strip ($Hc = 1{\cdot}2$, $\alpha = \frac{1}{2}\pi$). The numbers show the relative intensities along the isophotes.

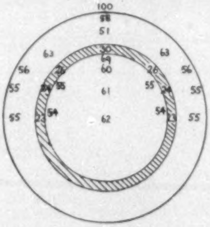


Figure 6. Flat-topped local zone $0{\cdot}6H \leqslant r \leqslant 0{\cdot}7H$, one twentieth of a fringe high, under test with a quarter-wave phase-retarding strip ($Hc = 1{\cdot}2$, $\alpha = \frac{1}{2}\pi$) set at focus. The numbers show the relative intensities seen in their immediate neighbourhoods.

From the results of the last two sections we may infer that when carried out with a slit-source of negligible width the phase-strip test can safely be used to figure an optical surface to an accuracy of one twentieth of a fringe. Below this limit, the systematic errors of the test begin to reduce its usefulness for the control of slow errors. Rapid local errors are shown with sufficient fidelity right up to the limit of sensitiveness of the test, which for such errors is of the order of one fiftieth of a fringe.

REFERENCES

Burch, C. R., 1934, *Mon. Not. R. Astr. Soc.*, **94**, 384.

Linfoot, E. H., 1946a, *Proc. Phys. Soc.*, **58**, 759; 1946b, *Proc. Roy. Soc.* A, **186**, 72; 1948, *Ibid.*, **193**, 248.

Zernike, F., 1934a, *Physica*, **1**, 689; 1934b, *Mon. Not. R. Astr. Soc.*, **94**, 377.

LETTERS TO THE EDITOR

On the Frequency Response of PbS Transistors

As described in a previous communication (Gebbie, Banbury and Hogarth 1950), some specimens of (deficit type) lead sulphide have been used successfully as crystal triodes, and voltage gains up to 60 have now been observed. It was of interest to compare the frequency response of such triodes with that of germanium transistors and to examine whether this response could be influenced by transverse magnetic fields as observed for n-type germanium (Bradner Brown 1949). Figure 1 shows response curves for one Ge and three PbS triodes, for 4 v. (1, 2, 3) and 6 v. (4) on the collector. It has been found that this frequency response of PbS triodes is dependent on the collector voltage. Thus, on a typical specimen for collector voltages of 1, 3, 6 and 8 volts, the corresponding frequencies at which voltage gain is reduced by 50% are 420, 460, 550 and 610 kc/s. Qualitatively this is expected from considerations of transit time of the charge carriers which pass from the emitter to the collector. Higher collector voltages cannot, at present, be used on PbS, but future developments in this direction would be expected to lead to higher frequency limits.

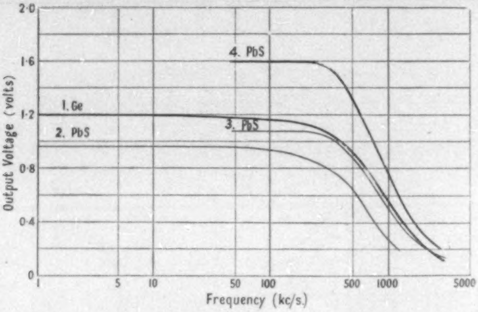


Figure 1. Frequency response of Ge and PbS transistors.

A magnetic field of about ±3,700 gauss was applied in the manner shown in Figure 2, and the corresponding values of voltage gain re-determined at various frequencies. The results (for 6 v. on the collector) are given in the following Table, together with similar measurements on Ge for purposes of comparison. It should be noted that these results are not directly comparable with those reported by Bradner Brown (1949), whose observations concerned *current gain*.

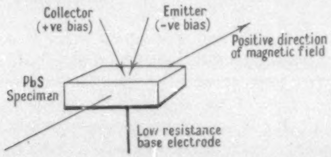


Figure 2. Direction of magnetic field relative to transistor electrodes.

Voltage Output for Fields of +3,700, 0, −3,700 gauss.

Frequency (kc/s.)	PbS (p-type): 0·05 v. input			Ge (n-type): 0·1 v. input		
	+3700	0	−3700	+3700	0	−3700
100	1·22	1·19	1·16	0·97	1·02	1·07
200	$0{\cdot}94_0$	$0{\cdot}92_0$	$0{\cdot}90_0$	0·96	1·01	1·07
300	$0{\cdot}68_5$	$0{\cdot}67_0$	$0{\cdot}65_5$	0·87	0·93	0·97
500	$0{\cdot}37_0$	$0{\cdot}36_0$	$0{\cdot}35_0$	0·68	0·73	0·76
700	$0{\cdot}25_0$	$0{\cdot}24_0$	$0{\cdot}23_0$	0·52	0·57	0·59
1000	$0{\cdot}16_5$	$0{\cdot}16_5$	$0{\cdot}16_0$	0·38	0·40	0·42
1500	$0{\cdot}12_5$	$0{\cdot}12_5$	$0{\cdot}12_0$	0·24	0·26	0·28
2000	—	—	—	0·17	0·18	0·20

According to the familiar transistor concepts, the magnetic field is expected to affect the average path length of the charge carriers. This length should be reduced by fields which tend to deflect the current flow from emitter to collector towards the surface. This would reduce the transit time which, together with the life-time of the carriers, limits the gain and its frequency response.

Although the effect of the magnetic field is small, it can be seen from its polarity that the charge carriers injected by the emitter into the PbS crystal must be electrons. The mechanism of this triode action is expected to be closely analogous to that of the Ge p-type transistor as reported by Pfann and Scaff (1949).

Thanks are due to Professor R. W. Ditchburn for placing research facilities at our disposal and to Dr. C. A. Hogarth for much helpful advice.

Department of Physics,
The University,
Reading.
20th April 1950.

P. C. Banbury.
H. K. Henisch.

Bradner Brown, C., 1949, *Phys. Rev.*, **76**, 1736.
Gebbie, H. A., Banbury, P. C., and Hogarth, C. A., 1950, *Proc. Phys. Soc.* B, **63**, 371.
Pfann, W. G., and Scaff, J. H., 1949, *Phys. Rev.*, **76**, 458 (A).

REVIEWS OF BOOKS

Medical X-Ray Protection up to Two Million Volts, National Bureau of Standards. Handbook 41. Pp. 41. (Washington: Superintendent of Documents, U.S. Department of Commerce, March 1949.) 15 *c*.

The present handbook supersedes a previous one of the National Bureau of Standards, last revised in 1936. It contains general sections, rules for specific applications in various kilovoltage regions and particularly details of electrical protection. The protective measures are based upon a permissible dosage-rate of 0·3 r/week in place of 0·5 r/week adopted in the recently issued Recommendations of the British X-Ray and Radium Protection Committee (October 1948), but it appears that the two conditions of measurement are different in the two cases, and over a wide range of radiations the numbers are essentially the same. While the recommendations are intended primarily for the protection of the radiation worker and not for the patient, the preface very rightly emphasizes the importance of caution and restraint in the use of x-rays for diagnosis and the treatment of non-malignant conditions. The genetic effects of radiation increase in importance as a larger fraction of the population is exposed to ionizing radiation, and diagnostic x-ray examinations become more and more frequent. Moreover, the employment of a large number of people in industries making use of ionizing radiations calls for, and is being given, serious thought. The present handbook merely hints at these problems and describes the consultative and advisory structure set up in the United States to deal with them.

Study of the recommendations shows that they are thorough and very detailed. A particularly useful appendix gives radiation attenuation curves in lead and concrete for beams of x-rays produced at voltages up to 2,000 kv. with worked examples illustrating practical calculations of protective requirements.

The handbook should be in use in all medical institutions using high voltage x-rays. It also contains much information which is appropriate to industrial organizations making use of these radiations.

W. V. M.

CONTENTS FOR SECTION A

PAGE
Mr. SURAJ N. GUPTA. Theory of Longitudinal Photons in Quantum Electrodynamics 681
Prof. M. H. L. PRYCE. On the Energy of the Alpha-Particles from Radioactive Nuclei 692
Dr. W. E. MOFFITT. Excited Electronic Levels in Conjugated Molecules—IV: Symmetrical Cyanine Dyes 700
Mr. J. EWLES and Dr. C. CURRY. Resolution and Analysis of Low Temperature Luminescent Spectra of Bi and Pb Activated Solids of Simple Crystal Structure 708
Dr. A. A. SABRY. Influence of Dipole–Dipole Coupling on the Dielectric Constant of Dipolar Substances 716
Dr. P. WRIGHT. The Effect of Occluded Hydrogen on the Electrical Resistance of Palladium 727
Dr. R. S. TEBBLE, Dr. I. C. SKIDMORE and Dr. W. D. CORNER. The Barkhausen Effect 739
Dr. K. H. STEWART. Domain Wall Movement in a Single Crystal 761
Prof. A. RUBINOWICZ. Sommerfeld's Polynomial Method Simplified 766
Dr. M. W. FEAST. Investigation of the Spectrum of the High Voltage Arc in Carbon Dioxide: the CO Flame Spectrum 772
Letters to the Editor:
Dr. F. D. S. BUTEMENT. Radioactive ^{165}Er 775
Mr. E. R. RAE, Mr. J. G. RUTHERGLEN and Mr. R. D. SMITH. Proton Capture Radiation of Fluorine-19 775
Prof. H. FRÖHLICH. Isotope Effect in Superconductivity 778
Dr. A. G. GAYDON and Dr. H. G. WOLFHARD. Excitation of Spectra in the Inner Cones of Flames 778
Corrigenda 780
Reviews of Books 780
Contents for Section B 783
Abstracts for Section B 783

ABSTRACTS FOR SECTION A

Theory of Longitudinal Photons in Quantum Electrodynamics, by SURAJ N. GUPTA.

ABSTRACT. The radiation field is quantized by introducing four types of photons—two transverse, one longitudinal, and one scalar. The scalar photons are treated by using an indefinite metric, and it is found necessary to modify the usual supplementary condition slightly. The present theory offers a justification for the symmetrical treatment of the four components of the electromagnetic potential, recently applied by a number of authors, and proves to be very convenient in applications. The results of physical interest, however, are the same as obtained from the ordinary formulation.

On the Energy of the Alpha-Particles from Radioactive Nuclei, by M. H. L. PRYCE.

ABSTRACT. When the energy of α-particles emitted by radioactive nuclei is compared with the energy calculated from the usual semi-empirical formula for nuclear binding energies, systematic variations ranging from 4·5 MeV. to −1·0 MeV. are found, which exhibit in a very clear way the special importance of configurations containing 82 protons and/or 126 neutrons. The general features are readily explained in terms of the shell model. The analysis can be used to predict unknown α-energies. Calculated α-energies are tabulated for the heavy nuclei ($Z \geqslant 78$) and for nuclei in the rare-earth region.

Excited Electronic Levels in Conjugated Molecules—IV: Symmetrical Cyanine Dyes, by W. E. MOFFITT.

ABSTRACT. Topics in the electronic theory of the colour of symmetrical cyanine dyes have been discussed by means of resonance formalism. Our conclusions for the formamidinium ion have been confirmed by use of the molecular orbital method. As a result of this analysis, certain modifications in the qualitative symbolism of colour theory have been put forward in the hope that the factors governing absorption may be more clearly understood.

Resolution and Analysis of Low Temperature Luminescent Spectra of Bi and Pb Activated Solids of Simple Crystal Structure, by J. EWLES and C. CURRY.

ABSTRACT. Experimental arrangements are described by which clearly resolved spectra have been obtained of the luminescence, at the temperature of liquid air, of seven different phosphors of simple crystal structure (cubic) and of mixed crystals. The main peaks in each case may be arranged in a band series. The vibration frequency in the lower state is nearly the same for the same host lattice with different activators. The energy of the electronic transition is determined both by the activator and the host crystal and there is a progression of decreasing energy in the series of simple crystals MgO, CaO, SrO, CaS, and in the order CaO, CaO–SrO with mixed crystals. The most probable transitions are from the lowest level of the upper state. There is evidence that self-excitation occurs and for the view that the luminescent centres are large.

Influence of Dipole–Dipole Coupling on the Dielectric Constant of Dipolar Substances, by A. A. SABRY.

ABSTRACT. Kirkwood's formula for dielectric constant, ϵ, gives it in terms of a correlation coefficient which depends on the probability that any two dipolar molecules have a certain configuration.

The main object of this paper is to calculate the configurational probability of the dipole orientations; this is dealt with in the first part of the paper by considering two dipoles and replacing the rest by a homogeneous dielectric. The dipolar molecules are regarded as point dipoles and the effect of other forces on the orientation of the molecules is neglected.

In the second part of the paper the calculation of the correlation coefficient appearing in Kirkwood's dielectric constant formula is carried out. This correlation coefficient is evaluated for water, and the value of ϵ then obtained from Kirkwood's formula is in fair agreement with the observed value.

The Effect of Occluded Hydrogen on the Electrical Resistance of Palladium, by P. WRIGHT.

ABSTRACT. Agreement between the results of theoretical and experimental investigations of the absorption of hydrogen by palladium indicates that, at least at higher temperatures, the hydrogen is contained in solid solution. Apparatus for observing the changes in electrical resistance of palladium during absorption and evolution of hydrogen is described, with which resistance–pressure isotherms from 75° to 150° c. were obtained for pressures increasing and decreasing between zero and atmospheric. Factors affecting the resistance of the Pd–H system are discussed and an explanation of the form of the isotherms is given; a decrease in resistance which is observed, when hydrogen is added to the alloy in the β phase, is attributed to completion of the d shells of palladium.

The Barkhausen Effect, by R. S. TEBBLE, I. C. SKIDMORE and W. D. CORNER.

ABSTRACT. An account is given of further experiments on the Barkhausen effect in cylindrical ferromagnetic specimens. The form of the decay of induction following a Barkhausen discontinuity has been investigated and there is reasonable agreement between the theory given and the experimental results. As a result of this investigation it has been possible to satisfy the different conditions which are necessary for the detection of the discontinuities so as to obtain (*a*) accurate reproduction of the form of the discontinuities and (*b*) maximum sensitivity in the counting experiments. The number and size of the discontinuities are measured, and an estimate is made of the contribution of the discontinuous processes to the total change in magnetization over the relevant part of the hysteresis curve. Discontinuous changes in magnetic moment greater than $0{\cdot}3 \times 10^{-6}$ E.M.U. (corresponding to a minimum volume of $0{\cdot}9 \times 10^{-10}$ cm^3 for iron) account for from 47% to 86% of the total change in magnetization. This is considered in relation to associated work on reversible changes, and the nature of a Barkhausen discontinuity is briefly discussed with reference to other investigations on domain structure.

Domain Wall Movement in a Single Crystal, by K. H. STEWART.

ABSTRACT. A single crystal of silicon–iron was shaped so as to encourage the formation of large ferromagnetic domains. Its magnetization curve was found to contain a large Barkhausen jump, indicating reversal of magnetization in a volume of the order of 1/10 cm^3. The rate of change of magnetization was found to be proportional to the excess of the applied magnetic field over a certain critical field, and could be made very slow. It is suggested that the rate of change is controlled by the braking effect of eddy currents on the movement of domain walls.

Sommerfeld's Polynomial Method Simplified, by A. RUBINOWICZ.

ABSTRACT. From two earlier papers by the author it follows that eigenfunctions obtainable by Summerfeld's polynomial method contain either a Riemann *P*-function or a degenerate form of it when the fundamental interval is either finite or infinite respectively. Comparing the differental equation for the *P*-function or its degenerate form with the differential equation which we get for *P* from the original differential equation of the eigenvalue problem, we obtain standard formulae for the constants determining the eigenvalues and eigenfunctions. The calculations necessary to give the complete solution of an eigenvalue problem which can be treated by Sommerfeld's polynomial method, are thus reduced to a minimum of diffentiations and simple algebraic operations. It is also possible, using appropriately chosen *P*-functions, to obtain the eigenfunctions directly in the required form. The method yields also eigenfunctions for eigenvalues in the continuous spectrum.

Investigation of the Spectrum of the High Voltage Arc in Carbon Dioxide: the CO Flame Spectrum, by M. W. FEAST.

ABSTRACT. The spectrum emitted by a high voltage arc in CO_2 at atmospheric pressure has been investigated in the region 2000 A. to 9000 A. It was found to consist, mainly, of a continuum and the ' CO flame bands '. A study of the spectrum at different dispersions and on photographic plates of various kinds shows that although the Schumann–Runge O_2 bands do form a part of the spectrum they are relatively weak compared with the bands attributed to CO_2. The excitation of the Schumann–Runge O_2 system in flames is discussed.